Antimetabolite biosynthesis in *Pseudomonas aeruginosa*

Von der Fakultät für Lebenswissenschaften

der Technischen Universität Carolo-Wilhelmina zu Braunschweig

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

genehmigte

Dissertation

von Yafei Xiao
aus Henan, China

1. Referent: Prof. Dr. Wulf Blankenfeldt
2. Referentin: Prof. Dr. Susanne Engelmann
eingereicht am: 22.08.2018
mündliche Prüfung (Disputation) am: 08. 11. 2018

Druckjahr 2018

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Tagungsbeiträge:

Yafei Xiao, Wulf Blankenfeldt. Antimetabolite biosynthesis in *Pseudomonas aeruginosa*. (Poster) 6th Annual Retreat, Brunsviga Braunschweig, Braunschweig, Germany, June 2015

Yafei Xiao, Wulf Blankenfeldt. Antimetabolite biosynthesis in *Pseudomonas aeruginosa*. (Poster) 8th International PhD Symposium of the Helmholtz International Graduate School, Helmholtz Centre for Infection Research, Braunschweig, Germany, December 2015

Yafei Xiao, Wulf Blankenfeldt. Antimetabolite biosynthesis in *Pseudomonas aeruginosa*. (Vortrag) 7th Annual Retreat, Goslar-Hahnenklee, Goslar, Germany, June 2016

Yafei Xiao, Wulf Blankenfeldt. Antimetabolite biosynthesis in *Pseudomonas aeruginosa*. (Poster) 9th International PhD Symposium of the Helmholtz International Graduate School, Helmholtz Centre for Infection Research, Braunschweig, Germany, December 2016

Yafei Xiao, Wulf Blankenfeldt. Antimetabolite biosynthesis in *Pseudomonas aeruginosa*. (Vortrag) 8th Annual Retreat, Goslar-Hahnenklee, Goslar, Germany, May 2017

Yafei Xiao, Wulf Blankenfeldt. Antimetabolite biosynthesis in *Pseudomonas aeruginosa*. (Poster) 4th Workshop of the Helmholtz Association “Cross Programme Activity Structural Biology”, Duesseldorf, Germany, November 2017

Yafei Xiao, Wulf Blankenfeldt. Antimetabolite biosynthesis in *Pseudomonas aeruginosa*. (Poster) 10th International PhD Symposium of the Helmholtz International Graduate School, Helmholtz Centre for Infection Research, Braunschweig, Germany, December 2017

Yafei Xiao, Jan Pippel, Wulf Blankenfeldt. Antimetabolite biosynthesis in *Pseudomonas aeruginosa*. (Poster) 26th Annual Meeting of the German Crystallographic Society (DGK), Essen, Germany, March 2018

Index

Abbreviations

Å	Angstrom (0.1 nm)
AMB	*L*-2-amino-4-methoxy-trans-3-butenoic acid, or methoxyvinylglycine
AVG	*L*-2-Amino-4-(2-aminoethoxy)-trans-3-butenoic acid
bp	Base pair
Da	Dalton
dCTP	Deoxycytidine triphosphate
DESY	Deutsches Elektronen-Synchrotron, Hamburg, Germany
DNA	Deoxyribonucleic acid
DTT	1, 4-dithiothreitol
E. coli	*Escherichia coli*
EDO_P8K	40% ethylene glycol and 20% polyethylene glycol 8000
ESRF	European Synchrotron Radiation Facility, Grenoble, France
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl] ethanesulfonic acid
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth
MR	Molecular replacement
MWCO	Molecular weight cut-off
NMR	Nuclear Magnetic Resonance
NOG	N-oxalyl glycine
OD_{600}	Optical density at 600 nm
PCR	Polymerase Chain Reaction
PEG	Polyethylene glycol
Phyre^2	Protein HomologY Recognition Engine V 2.0
r.m.s.d.	Root mean square deviation
Rpm	Revolutions per minute
SAM	S-adenosyl-methionine
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SeMet	Seleno-*L*-methionine

SLIC	Sequence and Ligation-Independent Cloning
SLS	Swiss Light Source, Paul Scherrer Institute, Villigen, Switzerland
SOB	Super optimal broth
TEV	Tobacco etch virus
Tm	Melting temperature
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol
UV	Ultraviolet

1. Summary

L-2-amino-4-methoxy-trans-3-butenoic acid (methoxyvinylglycine, or short as AMB) is a γ-substituted vinylglycine that was identified in 1972 as an inhibitor of pyridoxal 5'-phosphate (PLP)-dependent enzymes (Scannel *et al.*, 1972), resulting in antibiotic activity. Recent data indicate that the gene cluster *ambABCDE* is responsible for the AMB synthesis in *Pseudomonas aeruginosa* (Rojas Murcia *et al.*, 2015). Among the products of this gene cluster, AmbC and AmbD, two α-ketoglutarate-dependent dioxygenases are involved in the modification of *L*-glutamate as one substrate for AMB synthesis. AMB is then synthesized from one modified glutamate and two alanine moieties condensed to a tripeptide by the non-ribosomal peptide synthetases (NRPS) AmbB and AmbE, However, exact timing of the tailoring reactions remains unclear so far. Structural research was primarily focused on AmbC and AmbD. Additionally, the structure of unknown domain in NRPS AmbE needs to be determined to gain information on its function in AMB synthesis.

X-ray crystallography was mainly employed to obtain structures of the targeted proteins, combined with mass spectrometry (MS) experiments to provide a model for biosynthesis of AMB in *Pseudomonas aeruginosa*.

In this study, the structures of AmbD and AmbC 30-350, which was truncated based on predictions by Phyre^2 (Kelley *et al.*, 2015) and PSIPRED (Buchan *et al.*, 2013; Jones, 1999) have been determined. As the α-ketoglutarate-dependent dioxygenases, the complex structures of AmbD with α-ketoglutarate and AmbC 30-350 with α-ketoglutarate analog N-oxalylglycine (NOG) were also determined. Explorations on the other ligand which will be the specific intermediate synthesizing the final product AMB were carried out by co-crystallization and MS, while the progress was limited. By comparing these two complex structures with other available α-ketoglutarate-dependent dioxygenases (CarC, PDB: 4OJ8 (Chang *et al.*, 2014) and BBOX1, PDB: 3O2G, to be published) which contain the specific substrates, several potential substrate binding residues could be identified.

The NRPS AmbE contains a domain of unknown function whose structure could not be predicted. The specific sequence AmbE 20-420 was identified, and its structure

was determined then submitted to the PDBeFold (E. Krissinel & Henrick, 2004) and Dali server (Holm & Rosenstrom, 2010), suggesting it to be related to NRPS condensation domains. However, there is no conserved motif HHxxxDG that would be expected for condensation domains (Marahiel *et al.*, 1997; Rausch *et al.*, 2007; Stachelhaus *et al.*, 1998). Further, the two condensation domains within AmbB and AmbE are sufficient to synthesize a tripeptide product, such that a third condensation domain seems unnecessary.

2. Introduction

2.1 Antimetabolites and their function

Antimetabolites are chemicals that inhibit the function of metabolites, which are other chemicals that are part of the normal metabolism. A metabolite's antimetabolite can form a complex with the enzyme with which the metabolite normally reacts, but this complex cannot be converted into the normal reaction products. Nature long time ago has seized upon this phenomenon, possibly as an elegant way to control natural processes (N. A. de Bruyne, 1952).

Antimetabolites interfering with DNA production can be used in cancer treatment, as they inhibit cell division and tumor growth (Godefridus J. Peters, 2014; G. J. Peters *et al.*, 2000). Antimetabolites can also act as antibiotics that inhibit or reduce the activity of certain enzymes of pathogenic bacteria or viruses (Holladay, 2015).

2.2 The vinylic amino acids

Vinylic amino acids feature a C-C double bond directly attached to the α-carbon, along the side chain (Fig 2-1A). These vinylic amino acids can inhibit the activity of pyridoxal 5'-phosphate (PLP)-dependent enzymes by forming a more stable π-stacking system with PLP. Normally, as a cofactor, PLP binds to the ammonium group of a conserved lysine residue of an enzyme (Eliot & Kirsch, 2004). When the vinylic amino acid is introduced, the PLP can then bind the amino group of the vinylic amino acid, but because of the C-C double bond, the binding is much stronger than with the enzyme, resulting in the production of a stable vinylogous imidate, further leading to irreversible inhibition (Fig 2-2).

L-2-Amino-4-(2-aminoethoxy)-trans-3-butenoic acid (Aminoethoxyvinylglycine, short as AVG, Fig 2-1C) is biosynthesized in *Streptomyces* sp. NRRL5331 (Pruess *et al.*, 1974), and is an inhibitor of 1-aminocyclopropane-1-carboxylate (ACC) synthase that converts S-adenosyl methionine (SAM) into ACC. Since this reaction step plays a key role in plant ethylene biosynthesis, AVG is widely used as a fruit ripening inhibitor, and its inhibition has been characterized with enzymes from both tomato and apple (Capitani *et al.*, 2002; Huai *et al.*, 2001; Icekson & Apelbaum, 1983). A few crystal structures of ACC synthase in which PLP binds to AVG, have been determined and are shown in Fig 2-3.

2-amino-4-(2-amino-3-hydropropoxy)-trans-but-3-enoic acid (rhizobitoxine, Fig 2-1D), an anolog of AVG, is produced by symbiotic bacteria, such as *Bradyrhizobium japonicum* (Lowell D. Owens, 1972) and *Pseudomonas andropogonis* (Robin E.Mitchell, 1991), and its most important activity is inhibiting ACC synthase to regulate the ethylene biosynthesis (Yasuta *et al.*, 1999). Also rhizobitoxine inhibits β-cystathionase in the methionine biosynthesis pathway (K. Xiong, 1996).

2-amino-4-formylaminooxy-3-butenoic acid (4-formylaminooxyvinylglycine, short as FVG, Fig 2-1E) is produced by *Pseudomonas fluorescens* WH6 (Donald Armstrong, 2009). FVG irreversibly blocks germination of seeds from large number of grassy weed without significantly affecting the growth of mature plants (Halgren *et al.*, 2013). In addition to this herbicidal properties, it was found to inhibit the growth of *Erwinia amylovora*, the bacterial plant pathogen which causes firelight in orchard crops (Halgren *et al.*, 2011).

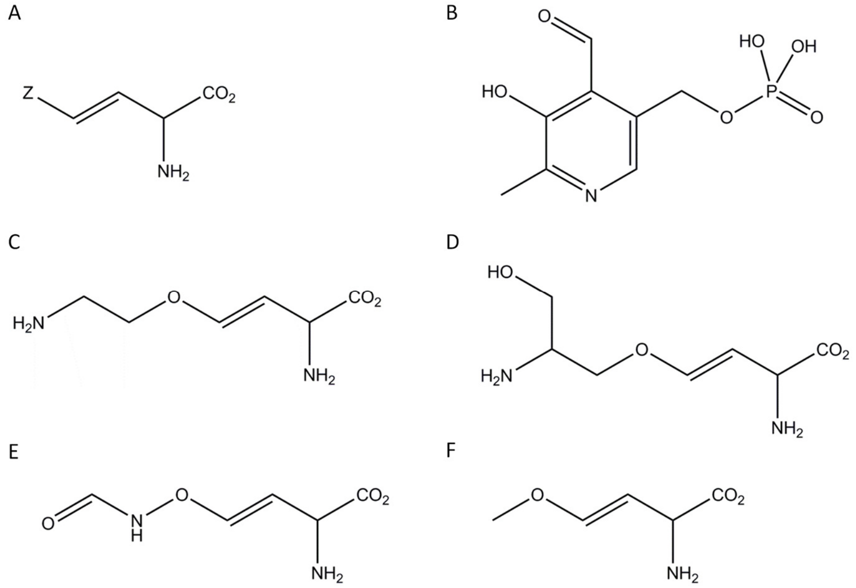


Fig 2-1: Chemical structures of vinylic amino acids (A), pyridoxal 5'-phosphate (PLP) (B), *L*-2-Amino-4-(2-aminoethoxy)-trans-3-butenoic acid (AVG) (C), 2-amino-4-(2-amino-3-

hydropropoxy)-trans-but-3-enoic acid (rhizobitoxine) (D), 2-amino-4-formylaminooxy-3-butenoic acid (FVG) (E) and *L*-2-amino-4-methoxy-trans-3-butenoic acid (AMB) (F).

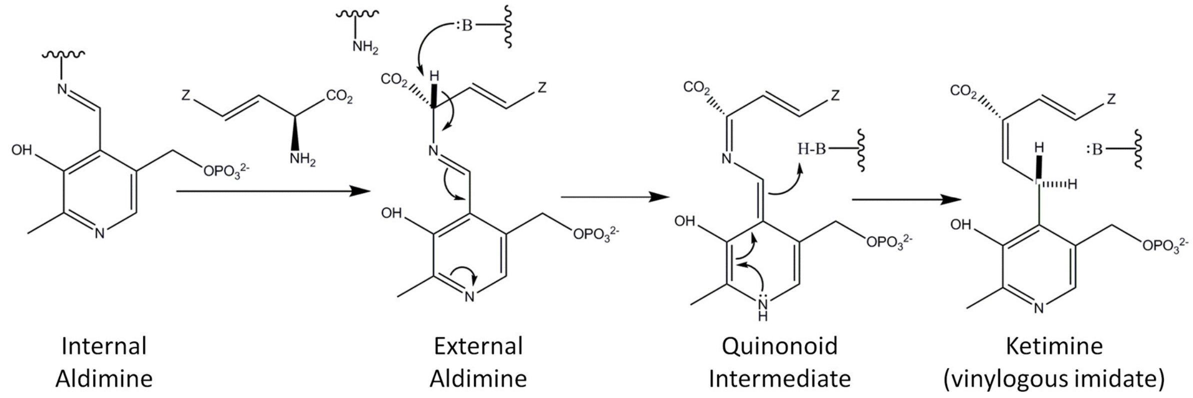


Fig 2-2: Mechanism of inhibition of vinylic amino acids to PLP-dependent enzymes.

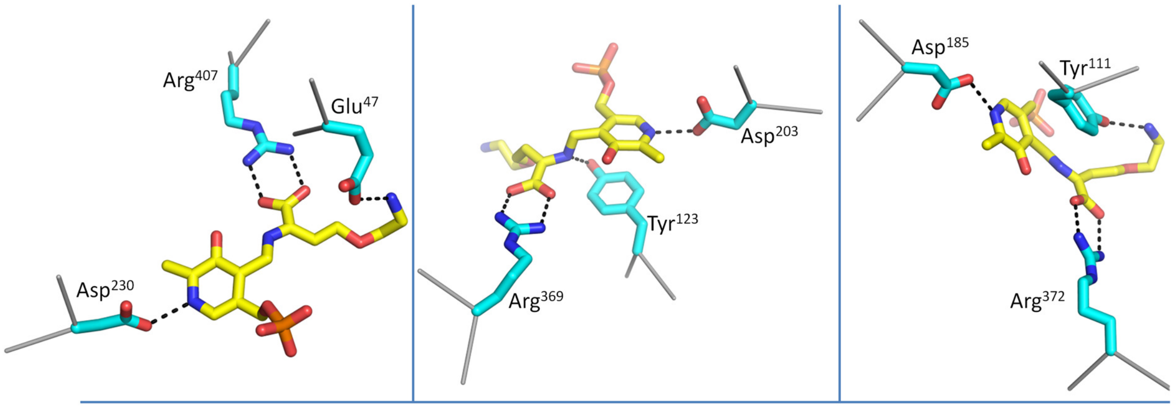


Fig 2-3: PLP-AVG adducts bound nearby the active site in three PLP-dependent enzymes of which PDB codes are 1M7Y, 1C7O and 1CL2 from left to right direction (Capitani *et al.*, 2002; Huai *et al.*, 2001; Icekson & Apelbaum, 1983). The carbons of PLP-AVG adducts are colored in yellow. Coordinated residues are colored in cyan, and nearby residues are shown as ribbon colored in grey.

2.3 *Pseudomonas aeruginosa* and AMB chemical characteristics

2.3.1 *Pseudomonas aeruginosa* is an opportunistic pathogen

The Gram-negative bacterium *Pseudomonas aeruginosa* favors moist areas and normally inhabits the soil and surfaces in aqueous environment. It is a facultative anaerobe (Schobert & Jahn, 2010) and can move by means of one or more polar flagella (van Ditmarsch *et al.*, 2013). At the same time, *Pseudomonas aeruginosa* is an opportunistic pathogen and can cause infections in plants (Rahme *et al.*, 1995; Rahme *et al.*, 1997; Walker *et al.*, 2004) (Fig 2-4) and animals (D'Argenio *et al.*, 2001; Mahajan-Miklos *et al.*, 1999; Miyata *et al.*, 2003; Rahme *et al.*, 2000), including humans. Especially persons who are immuno-compromised, such as patients with neutropenia, severe burns, cancer or HIV infections or patients with surgical wounds and organ transplantations (Lyczak *et al.*, 2000) are prone to get acute life-threatening infections resulting in bacteremia, urinary tract infections and pneumonia. *Pseudomonas aeruginosa* can also provoke chronic and persistent infections, especially in patients suffering from cystic fibrosis and chronic obstructive pulmonary disease (Arancibia *et al.*, 2002; Shaan L. Gellatly, 2013; Williams *et al.*, 2010). The bacterium can colonize on and in medical equipment, including catheters, resulting in cross-infections in hospital and clinics where the sanitary conditions are not sufficient. In the United States, *Pseudomonas aeruginosa* is among the most common hospital pathogens and is the second most common pathogen isolated from patients with ventilator-associated pneumonia (Hidron *et al.*, 2008). Treatment of *Pseudomonas aeruginosa* infections is compromised by its resistance to a large range of antibiotics (Breidenstein *et al.*, 2011; Cornelis, 2008; Jacoby, 2009; Poole, 2004; Schweizer, 2003; Y. F. Zhang *et al.*, 2017). Gain of additional resistances can be observed after unsuccessful treatment (Fernandez *et al.*, 2011).

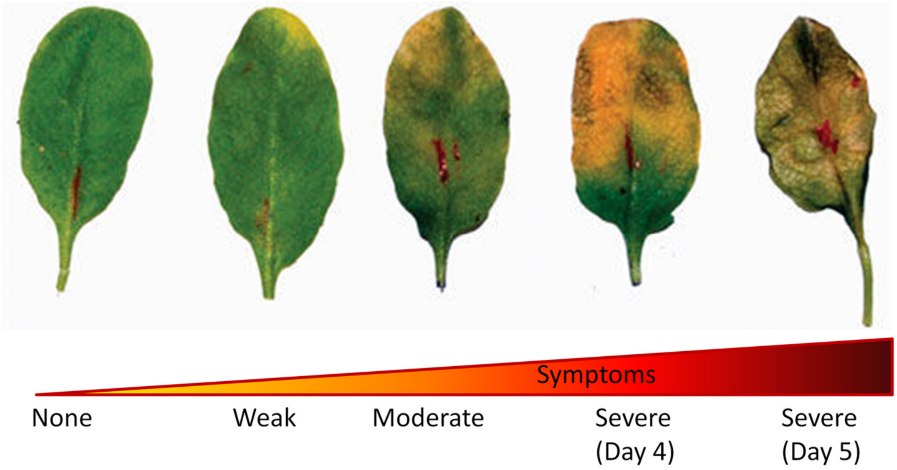


Fig 2-4: *Pseudomonas aeruginosa* infected *Arabidopsis* leaves. The figure is adapted from Starkey and Rahme (2009). The leaves show a range of symptoms from mild chlorosis to severe rotting.

Pseudomonas aeruginosa expresses virulence factors to support its growth and reproduction and to inhibit other species growth. Pyocyanin is one of these factors which can cause death of *Caenorhabditis elegans* by oxidative stress. Exotoxin is another virulence factor which can inhibit eukaryotic elongation factor-2 by ADP-ribosylation (Yates & Merrill, 2004) , thereby inhibiting protein synthesis.

2.3.2 AMB can inhibit bacterial growth

One of the many toxins produced by *Pseudomonas aeruginosa* is *L*-2-amino-4-methoxy-trans-3-butenoic acid (methoxyvinylglycine, or short AMB, Fig 2-1F), which was first found in 1972. This toxin can reversibly inhibit the growth of *Bacillus* sp (Scannel *et al.*, 1972) and *Escherichia coli* (Sahm *et al.*, 1973). AMB belongs to γ-substituted vinylglycine which can inhibit the activity of pyridoxal 5’-phosphate (PLP)-dependent enzymes (Huai *et al.*, 2001; Percudani & Peracchi, 2003; Rando, 1974; Rando *et al.*, 1976). As a potential antibiotic, more and more research focuses on AMB. It was reported to inhibit aspartate transaminase in pigs (Rando, 1974; Rando *et al.*, 1976) and rat hepatocytes (Cornell *et al.*, 1984; S. B. Smith & Freedland, 1981), tryptophan synthase in *E. coli* (Miles, 1975), δ-aminolevulinic acid synthase in rats (Dashman & Kamm, 1979), and serine hydroxylmethyl transferase in Walker carcinoma (Tisdale, 1980). At the meantime AMB, as a methionine antimetabolite, can inhibit *L*-methionine tRNA aminoacylation (Mattoo *et al.*, 1979).

By using an *Acanthamoeba castellanii* cell model, the importance of AMB as a virulence factor produced by *Pseudomonas aeruginosa* was evaluated (Lee *et al.*, 2012). It was found that the contribution of AMB to the virulence of *Pseudomonas aeruginosa* was not strong, because quite high concentrations were needed for inducing cyst formation. AMB can also inhibit the growth of important plant pathogens such as *Erwinia amylovora* (X. Lee *et al.*, 2013), which may make it possible to apply AMB in agriculture, similar to AVG, which is commercially available as ReTain® to regulate the ripening of fruits.

2.4 The gene cluster *ambABCDE* and protein analysis

2.4.1 Gene analysis and function prediction

In 2013, the *Pseudomonas aeruginosa* gene cluster *ambABCDE* was reported to be responsible for synthesis of IQS, which is a quorum sensing signal molecule (J. Lee *et al.*, 2013). Two years later, however, by a reverse genetics approach using transposon mutagenesis, the gene cluster *ambABCDE* was found to be rather responsible for the biosynthesis of AMB (X. Lee *et al.*, 2013; Lee *et al.*, 2010; Rojas Murcia *et al.*, 2015). There are total two transcriptional units in *ambABCDE*: *ambA* encodes a putative lysine exporter transmembrane protein and may thus control the secretion of AMB. The *ambBCDE* genes encode four different proteins. AmbC and AmbD are predicted to be iron (II)/α-ketoglutarate-dependent dioxygenases. AmbB and AmbE are both non-ribosomal peptide synthetases (NRPSs) of which the modular structures are shown in Fig 2-5. They have domains for adenylation (A), thiolation (T), condensation (C), methylation (MT) and thioester cleavage (TE). Special for AmbE, it has an uncharacteristic domain at its N-terminal, and an atypical C domain which is indicated as C*. Analysis of the C* domain suggests an additional function, as it is not similar to previously known epimerization, heterocyclization, or starter C domains. This C* domain rather groups with another atypical C domain found in the McyA protein from *Microcystis aeruginosa* (Tillett *et al.*, 2000).

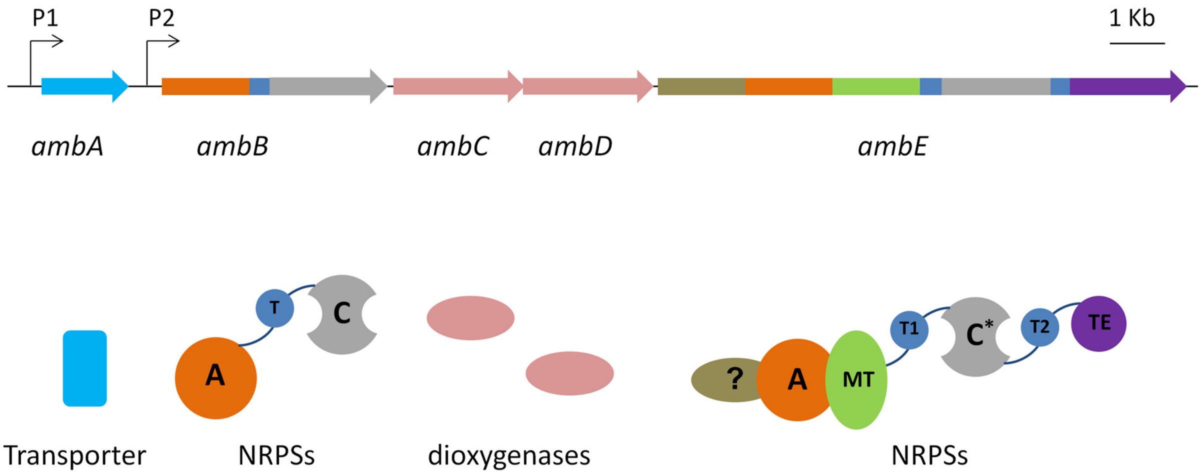


Fig 2-5: Organization of gene cluster *ambABCDE* in the *Pseudomonas aeruginosa* strain PAO1, and the prediction of proteins and functional domains.

Lee *et al.* analyzed AMB synthesis *in vitro* by mass spectrometry (MS) of enzyme-attached substrates and pathway intermediates. They identified the building blocks of AMB synthesis and proposed a model of how the AMB precursor tripeptide is assembled.

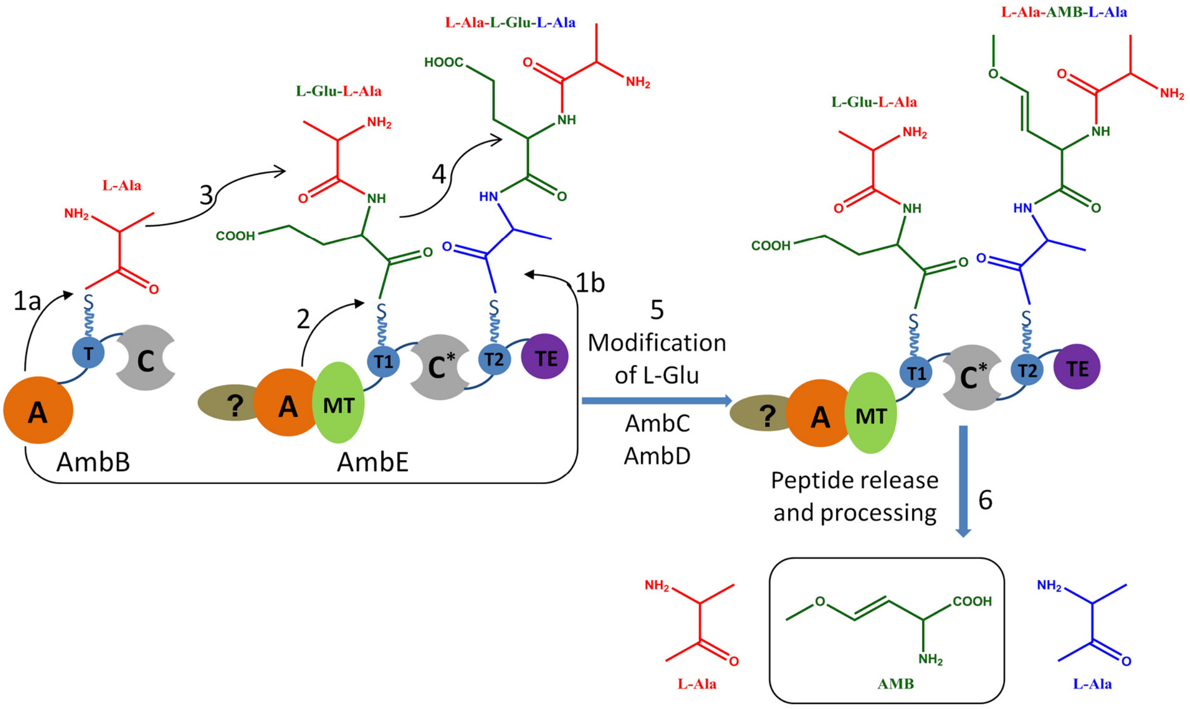


Fig 2-6: Proposed model for AMB biosynthesis with indication of tethered molecules. Modified figure from (Rojas Murcia *et al.*, 2015)

The model is illustrated in Fig 2-6. The A domain of AmbB activates two *L*-Ala (residues shown in red and blue) and loads them onto its own T domain (step 1a) and onto the T2 domain of AmbE (step 1b). The A domain of AmbE activates *L*-Glu (residues shown in olive) and loads it onto its T1 domain (step 2). Then, the C domain of AmbB would condense *L*-Ala with *L*-Glu onto the T1 domain of AmbE forming the dipeptide *L*-Ala-*L*-Glu (step 3). This dipeptide would subsequently be further condensed by the C* domain with *L*-Ala onto the T2 domain of AmbE to obtain the tripeptide *L*-Ala-*L*-Glu-*L*-Ala (step 4). Modification of the glutamate would then result in *L*-Ala-AMB-*L*-Ala (step 5). This peptide would be released by the TE domain of AmbE. Removal of the two flanking *L*-Ala residues would finally yield AMB (step 6).

By now it is unclear how *L*-glutamate moiety is modified and how AmbC, AmbD and the unknown domain of AmbE are involved.

2.4.2 α-ketoglutarate-dependent dioxygenases

2.4.2.1 Overview

Oxygenase was discovered in 1955 (H. S. Mason, 1955; Osamu Hayaishi, 1955) as an enzyme that can oxidize a substrate by transferring oxygen from O_2 to it. Oxygenase can be divided to two types: monooxygenase and dioxygenase. The former will transfer one oxygen atom of an O_2 molecule to the substrate and will incorporate the other into a water molecule. The latter will incorporate both oxygen atoms into products of the reaction (Bugg, 2003).

Most dioxygenase enzymes require a metal cofactor, which is most often iron (II) or iron (III) (Que & Ho, 1996). Normally these enzymes need an α-keto acid as a co-substrate, which is mostly the α-ketoglutarate (Costas M, 2004; Hausinger, 2004; A. G. P. a. M. D. Lloyd, 2000; Schofield & Zhang, 1999). Hydroxylation is the most common reaction carried out by these enzymes, and they can also perform many other interesting reactions: halogenation, oxidative ring formation, ring expansion, epimerization and desaturation (Purpero & Moran, 2007).

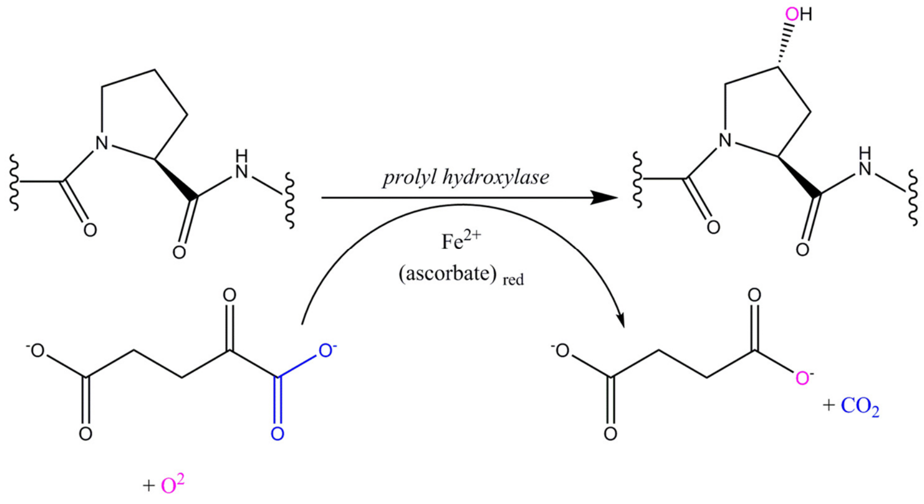


Fig 2-7: Reaction scheme for prolyl hydroxylase. This hydroxylation is catalyzed by ferrous iron. Co-substrate α-ketoglutarate is decarboxylated to yield one carbon dioxide (blue). The carboxyl group is substituted by an oxygen atom from molecular dioxygen (magenta), producing succinate. The other oxygen atom is incorporated into a prolyl residue, yielding 4-hydroxy-prolyl. The scheme was modified from Figure 24 in Bugg (2003)

The first identified α-ketoglutarate-dependent dioxygenase is prolyl hydroxylase (Hutton *et al.*, 1966), which catalyses the hydroxylation of prolyl moieties in collagen to 4-hydroxy-prolyl. One oxygen atom is transferred to the hydroxylated product, and the other one to α-ketoglutarate, generating succinate and CO_2 (Fig 2-7) (Cardinale *et al.*, 1971).

2.4.2.2 Structural and mechanistic analysis

The core structures of α-ketoglutarate-dependent enzymes are characterized by a “jelly roll” fold comprising eight anti-parallel β-strands forming two opposing β-sheets (Hewitson *et al.*, 2005; Roach *et al.*,1995). Other secondary structural elements and flexible loops surround and presumably stabilize these core sheets (Fig 2-8). These insertions and extensions to the “jelly roll” fold are different in each protein. They define substrate specificity and establish protein contacts. Enzymes of this family can be monomers, dimers, trimers, tetramers and hexamers (Clifton *et al.*, 2003; Clifton *et al.*, 2001; Dann *et al.*, 2002; Elkins *et al.*, 2003; M. D. Lloyd *et al.*, 1999; Muller *et al.*, 2004; O'Brien *et al.*, 2003; Wilmouth *et al.*, 2002; Z. Zhang *et al.*, 2000).

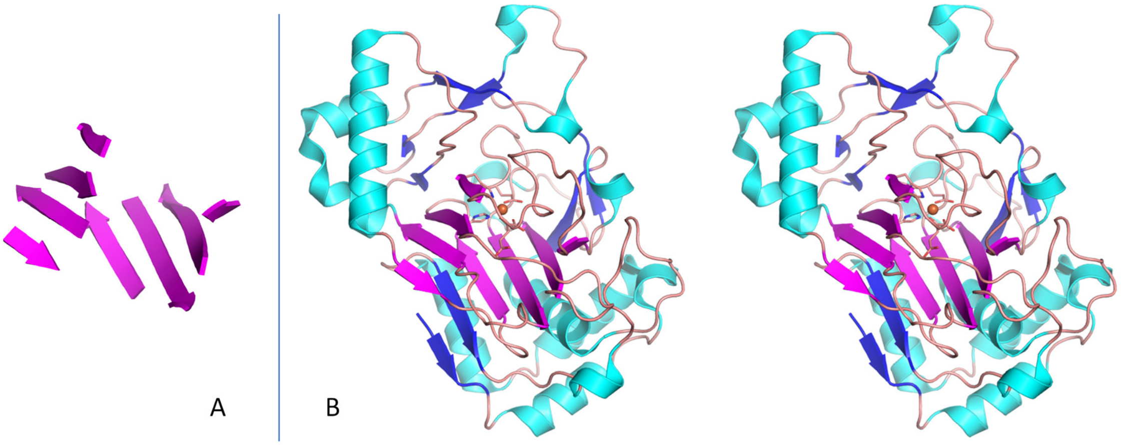


Fig 2-8: Example of the jelly roll fold found in Clavaminate synthase, one Fe (II)/α-ketoglutarate-dependent dioxygenases (PDB 1DS1, Z. Zhang *et al.*, 2000). A: Eight anti-

parallel β-strands combining to form the core of enzymes of this family. B: Overall stereo view of the Clavaminate synthase structure. Additional secondary structures are shown at the N and C terminus. This enzyme contains insertions and extensions to this jelly roll fold, which influence contacts to other proteins and substrate binding. α-helices, β-sheets and loops are colored in cyan, magenta and salmon, respectively. β-sheets that do not belong to the jelly roll are colored in blue.

The active sites of α-ketoglutarate-dependent dioxygenases have similar arrangements of three conserved residues, namely two histidines and one acid amino acid (aspartate or glutamate) (Hegg & Que, 1997). They coordinate the central metal ion through an octahedral arrangement. Two of three vacant binding sites around the metal are occupied by a molecule of the co-substrate α-ketoglutarate (Fig 2-9). The sixth site can be coordinated with a dioxygen molecule, however, in most cases this coordination is not possible before the specific substrate has bound (Costas *et al.*, 2004; Elkins *et al.*, 2002; Zhou *et al.*, 2001). After all substrates are recruited to their positions the reaction commences. A proposed reaction mechanism is shown in Fig 2-10.

In some cases, the active site can be closed off by a flexible loop or one secondary structure element (Purpero & Moran, 2007) to prevent the release of reactive oxygen species (ROS), which could cause serious damage on proteins, DNA, RNA and lipids (Simon *et al.*, 1981; Valko *et al.*, 2006).

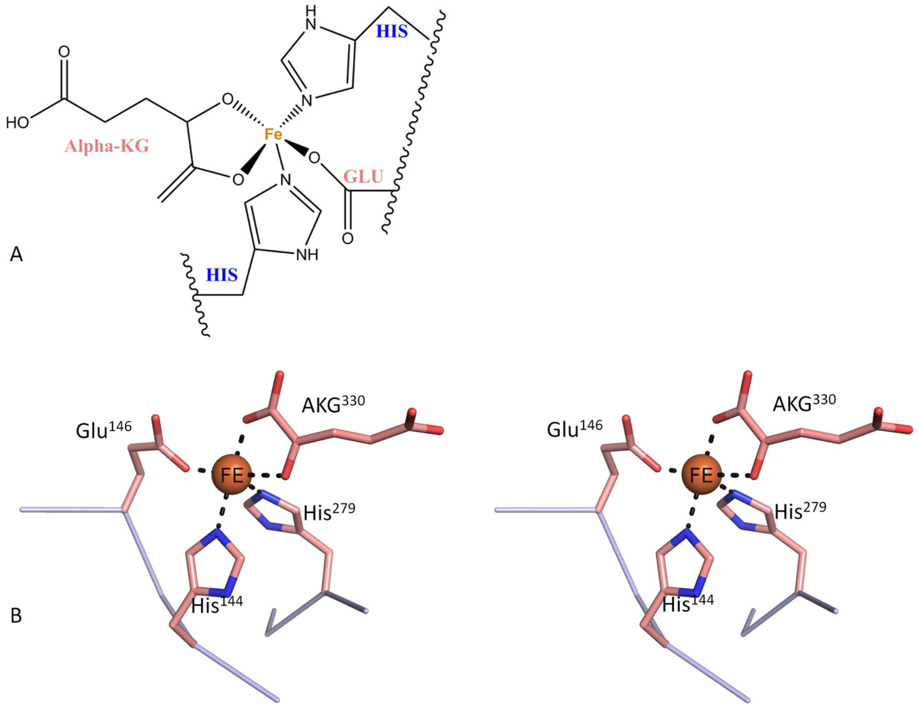


Fig 2-9: Active site structure of Clavaminate synthase, an α-ketoglutarate-dependent dioxygenase (PDB 1DS1, Z. Zhang *et al.*, 2000). A: Scheme of the active site. B: Stereo view of the active site from Clavaminate synthase. The His-X-Glu-X$_n$-His motif and the co-substrate α-ketoglutarate (AKG) are shown. They coordinate with the central metal iron in an octahedral arrangement. The sixth coordination site will be occupied by a dioxygen molecule after the substrate is introduced.

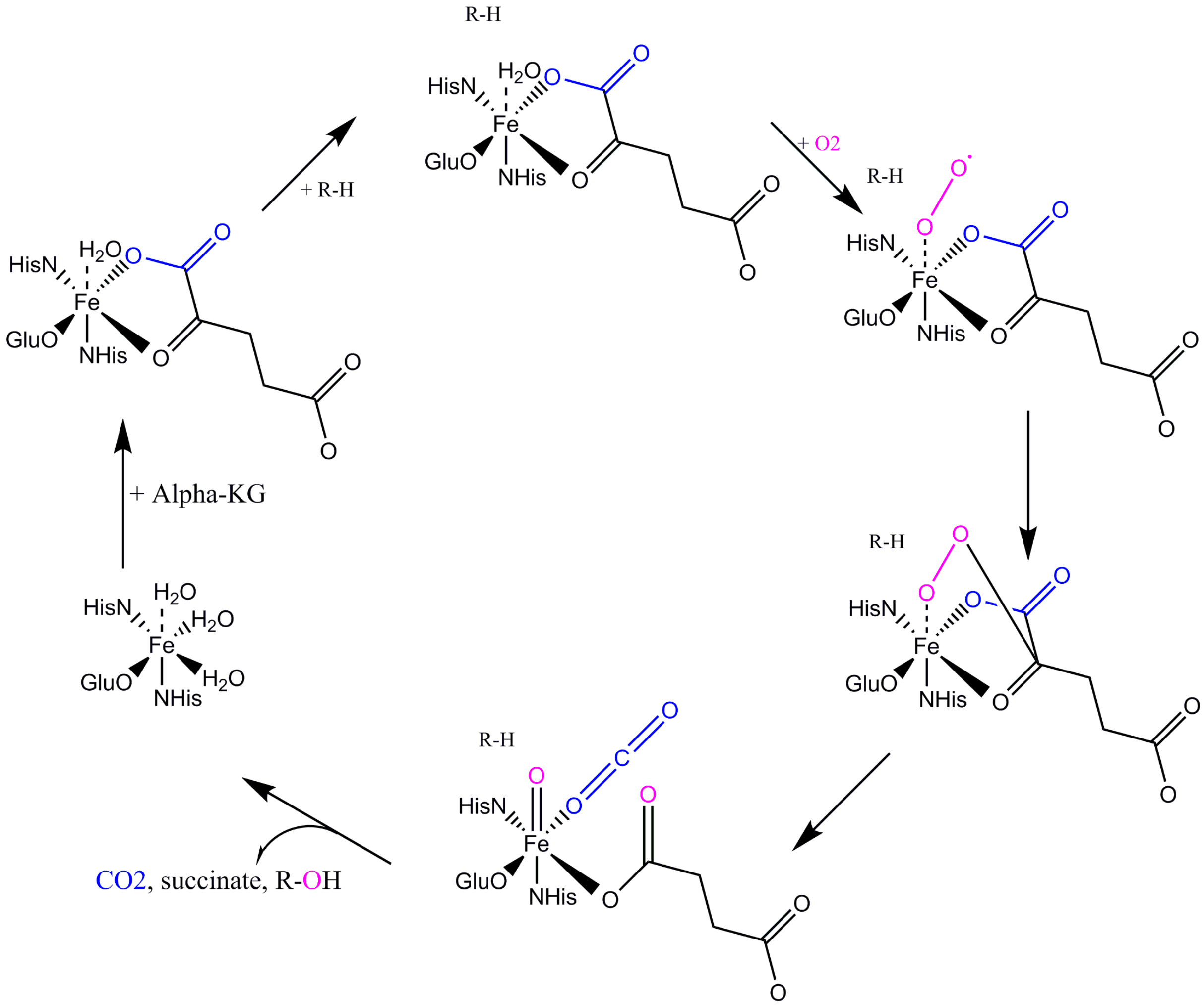


Fig 2-10: Proposed catalytic mechanism of α-ketoglutarate-dependent dioxygenases. The figure is based on Figure 26 from Bugg (2003). The same colors as in Fig 2-7 were used.

2.4.3 Non-ribosomal peptide synthesis (NRPS) procedures

2.4.3.1 Overview

The ribosome is the main molecular machine producing complex polypeptides or proteins. Bacteria (Byford *et al.*, 1997; Zhu *et al.*, 1998) and fungi (Bushley *et al.*, 2013; Correia *et al.*, 2003) apply an additional pathway to synthesize short peptides that uses non-ribosomal peptide synthetases (NRPSs). These non-ribosomal peptides (NRP) comprise a wide range of biological and pharmacological secondary metabolites, including antibiotics as well as toxins, siderophores and pigments (Vining, 1990). The size of NRPs can range from 2 to 48 amino acid residues (Kleinkauf & Von Dohren, 1996). In addition to amino acids, more than 500 other monomers such as non-proteinogenic and N-methylated amino acids, *D*-amino acids, fatty acids and α-hydroxy acids can be taken as building blocks (Caboche *et al.*, 2010).

2.4.3.2 The structure of non-ribosomal peptide synthetases (NRPSs)

Non-ribosomal peptide synthetases (NRPSs) are multidomain enzymes which act as assembly lines catalyzing the biosynthesis of complex natural products. Each NRPS is comprised of one or several modules to perform various reactions (Hahn & Stachelhaus, 2004). Within a single module, three domains are ubiquitous and essential for peptide elongation. They are (i) an ATP-dependent adenylation (A) domain to select and activate the specific amino acid or alternative building block; (ii) an thiolation (T, also called peptidyl carrier protein, PCP) domain which bears a 4'-phosphopantetheine (Ppant) group in its holoform to carry the activated amino acid or intermediate of non-ribosomal peptide and presents it to the catalytic site (Gevers *et al.*, 1969; Roskoski, Gevers*, et al.*, 1970; Roskoski, Kleinkauf*, et al.*, 1970; Stoll *et al.*, 1970); and (iii) a condensation (C) domain to elongate the upstream and downstream intermediates by forming a peptide bonds (Kleinkauf *et al.*, 1969). Additionally, methyltransferase, epimerase and cyclase domains can be present to modify the intermediate or substrate (Balibar *et al.*, 2005; Gaudelli & Townsend, 2014; Van de Poel & Van Der Straeten, 2014; Velkov & Lawen, 2003). At last, a terminal thioesterase (TE) domain will release the final product from the NRPS reaction platform (Gaudelli & Townsend, 2014; Puk *et al.*, 2004; Puk *et al.*, 2002) (Fig 2-11).

Generally, one NRP is synthesized by several cooperating NRPSs, which can interact via communication (COM) domains locating at the N- and C-termini (Hahn & Stachelhaus, 2004, 2006). Tailoring enzymes introduce further modifications and decorations to the NRP (Fridman *et al.*, 2007; Fu *et al.*, 2005; Galm *et al.*, 2004; Losey *et al.*, 2002; van Wageningen *et al.*, 1998; Xu *et al.*, 2004).

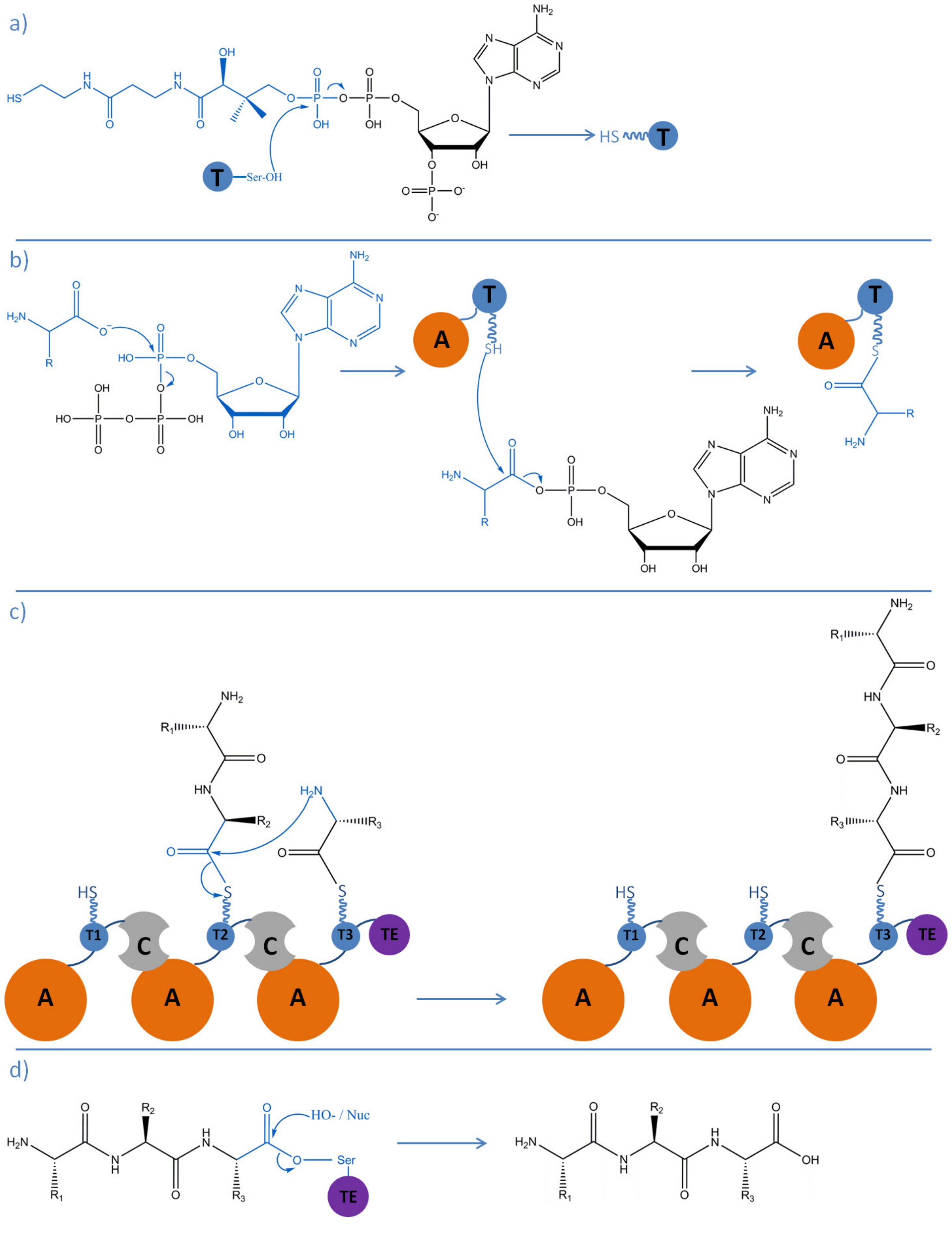


Fig 2-11: Reactions catalyzed by the NRPSs and their domain arrangement. a) The Ppant arm from coenzyme A is transferred to a conserved serine residue in the T domain to

obtain a holo state, which is essential for building blocks binding. b) The ATP-dependent A domain selects and activates amino acids, and catalyzes the formation of an aminoacyl thioester with the Ppant cofactor of the holo-T domain. c) Peptide elongation catalyzed by the C domain. A peptide bond forms between the monomer (acceptor) and the nascent peptide (donor), and results in translocation of the new peptide chain to the acceptor T domain. d) Release of final oligopeptide by the TE domain. First, the oligopeptide attached to last T domain is transferred to a conserved serine residue in the TE domain. Then hydrolysis or intramolecular attack of a nucleophilic moiety produces a linear or cyclic product, respectively. The moieties involved in product reactions are colored blue, and Nuc means nucleophile.

2.5 Research objectives

As an oxyvinylglycine, AMB is the first one generated from *L*-Glu. In other cases, AVG, rhizobitoxine and FVG, the initial substrate or amino acid appears to be homoserine. Additionally, these three compounds are not suggested to be synthesized by a thiotemplate mechanism, or processed from a precursor peptide.

The biosynthesis of AMB is under control of the gene cluster *ambABCDE*. Among the products of this gene cluster, proteins AmbC and AmbD are both predicted as α-ketoglutarate-dependent dioxygenases, which probably hydroxylate the *L*-Glu moiety. As tailoring enzymes, the determination of their structures will provide insight on specific substrates or intermediates which are unclear by now. Proteins AmbB and AmbE are NRPSs, which condense one glutamate and two alanine moieties to a tripeptide. Domain analysis provides some information on these huge proteins, except for an N-terminal area of AmbE. The structure of this uncharacterized domain should also be determined, combining with mass spectrometry (MS) experiments to unravel its exact function and timing points of the reactions and modifications of the *L*-Glu moiety.

All in all, X-ray crystallography and MS experiments are designed and performed to develop a biosynthesis model on the molecular level to explain how AMB is synthesized step by step.

3. Materials and methods

For this project, methods of molecular biology, chemical biology and protein crystallography were mainly applied. All the materials and methods are similar for different proteins, so only the commonly used ones are shown below.

3.1 Materials

3.1.1 E. coli strains

Table 3-1 strains

Strains	Main characters	Antibiotics Resistance
OmniMAX^{TM}2	Plasmids amplification	No
XL-1 Blue	Plasmids amplification	Tetracycline
DH 5α	Plasmids amplification	No
BL21 (DE3)	Protein over expression	No
Tuner (DE3)	Protein over expression	No
Rosetta2 (DE3) pLysS	Protein (using rare genetic codes) over expression	Chloramphenicol

3.1.2 Plasmids

Table 3-2 plasmids

Plasmids	Main genetic elements
pOPINB	T7 promoter, His tag with PreScission cleavage site, kanamycin resistance
pET27b	T7 promoter, His tag, kanamycin resistance
pET28a	T7 promoter, His tag with thrombin cleavage site, kanamycin resistance

3.1.3 Enzymes and kits

Table 3-3 enzymes and kits

Enzymes and kits	Vendor
KpnI-HF/ HindIII-HF	NEB
Calf Intestinal Phosphatase (CIP)	NEB
Phusion Flash Mix	Thermofisher
DpnI	NEB
T4 DNA Polymerase	Thermofisher
Recombinase A	Epicentre
DNaseI	Roche
Lysozyme	Fluka
Thrombin	Serva
QIAquick PCR Purification Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
QIAprep Spin Miniprep Kit	Qiagen

3.2 Main Instruments

Table 3-4 main instruments

Instruments (model)	Vendor
Äkta purifier	GE Healthcare Life Sciences
Clean bench (Herasafe KS 12)	Thermo Scientific
Electrophoresis Cell (Tetra Vertical)	Bio Rad
Homogenisator (EmulsiFlex-C3)	Avestin
Gel image system (Gel Logic 212)	Kodak
NanoDrop 2000	Thermo Scientific
PCR Thermocycler (peqSTAR)	VWR
Centrifuge (Micro Star 17R)	VWR
Ultracentrifuge (Sorvall RC 6+ centrifuge)	Thermo Scientific

3.3 Methods

3.3.1 Bioinformatic analysis

(1) Sequences: All gene sequences used in this project were retrieved from the website: http://pseudomonas.com/

(2) Basic property analysis of proteins: Molecular weight, pI and extinction coefficients were calculated by ProtParam on the website https://web.expasy.org/protparam/

(3) Protein domain analysis: Functional domain analyses were performed by Smart on the website http://smart.embl-heidelberg.de

(4) Homologous alignment: Homologous protein sequences were obtained by Blast on NCBI, and the ClustalW method was applied to perform sequence alignment.

(5) Prediction of secondary structure and disorder tendency: PSIPRED was applied on the website http://bioinf.cs.ucl.ac.uk/psipred

(6) Three-dimensional modeling: Homologous protein structures were retrieved from the PDB website, https://www.rcsb.org/. Structures were modeled by using Phyre2, http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index

(7) Prediction of protein folding tendency: The FoldIndex server was used, https://fold.weizmann.ac.il/fldbin/findex

3.3.2 Gene clone and plasmids construction

Polymerase chain reaction (PCR) was used to amplify target genes. By designing different primers with different restriction enzyme sites, targeted gene sequences were cloned. In this project, two different protein constructs were prepared by using Sequence and Ligation-Independent Cloning (SLIC). More details about this method are shown below.

At the beginning, double digested vector is prepared with restriction enzymes. According to the published method (Li & Elledge, 2007), primers containing tails of 20-

25 bps overlapping with the vector ends were designed. Also the Tm of full-length primers should be above 68 °C, and the Tm of the part of the primer matching the target sequence should be around 60 °C. In this project, two target protein regions, which were AmbC, amino acids 30-350, and AmbE, amino acids 20-420 were fully studied. Both coding sequences were cloned into pOPINB. Here the construct of pOPINB *AmbE 20-420* was shown as an example to explain the primer design.

First, the restriction enzymes KpnI (cleavage site at 5133 bp) and HindIII (5466 bp) were chosen to digest the vector pOPINB (fig 3-1). Consequently the primers were designed containing these two cleavage sites respectively. As shown in the table 3-5, the primers consisted of three parts: sequence of a vector end, restriction enzyme site and sequence of insert fragment. Because the HindIII cleavage site contains two thymidine deoxyribonucleotides (T), one adenine deoxyribonucleotide (A) was added to the end of complement sequence of insert fragment to compose a stop code (TAA) on the reverse primer. Both Tm of insert fragments are around 60 °C, and that of full-length primers are higher than 70 °C.

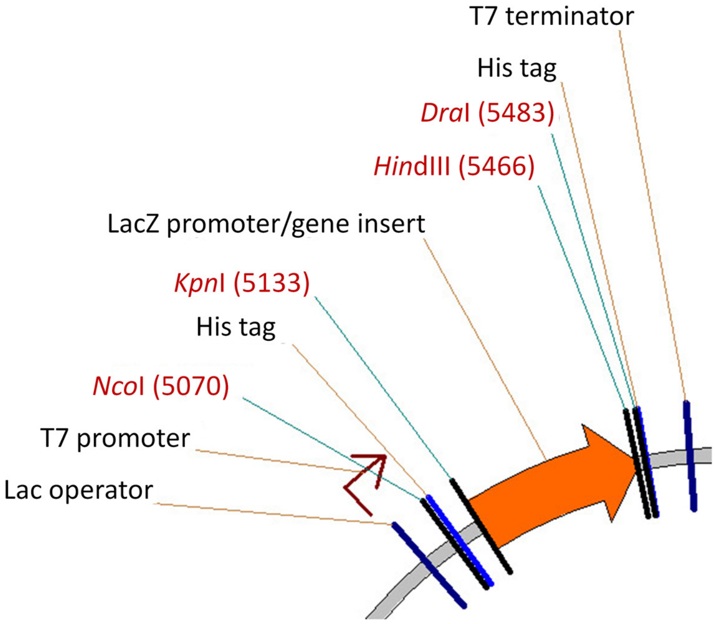


Figure 3-1: Part of pOPINB vector. The multiple cloning site area is shown. The KpnI and Hind III sites were chosen for cloning.

Table 3-5

F:	ctggaagttctgtttcag	GGTACC	gaaggattcccgctgtctccctt	47 bps
	Vector seq.	KpnI	target seq. (23 bps, Tm 62 °C)	Tm total 74 °C
R:	ggtgtttaaatggtctaga	AAGCTT	aagtggcagttgcctgtcggc	46 bps
	Vector seq.	HindIII	target seq. (20 bps, Tm 64 °C)	Tm total 71 °C

With these primers and template containing the full-length *ambE* gene, the PCR was performed to amplify the insert fragment with Phusion Flash Master Mix. See table 3-6 for ingredients and concentrations.

Table 3-6

PCR (one well)		Stock concentration	Final concentration
[µl]			
1	Template	100 ng/µl	3.3 ng/µl
4	Mixture of primers	2 µM	250 nM
15	Phusion Flash Mix	2 ×	1×
6	Betain	5 M	1 M
4	H_2O (sterilized)		

After mixing all the ingredients above, 30 µl reaction solutions were obtained. Then the PCR was performed for 30 cycles as shown in table 3-7.

Table 3-7

Steps	Temperature [°C]	Time [s]
Initial denaturation	98	300
Denaturation	98	10
Annealing	Around 55 °C (5 °C lower than Tm of insert fragment at least)	10
Elongation	72	70 (Elongation efficiency is 1 Kbp/45 s)

Last elongation	72	600
Store	4	∞

When this PCR finished, 1 µl DpnI enzyme and respective buffer were added to digest the template DNA. The amplified insert fragments were extracted by using PCR-purification kit (Qiagen), and agarose gel was applied to check the products purity and the final concentration was calculated by Nanodrop2000.

Digested vector and purified insert fragments were treated with T4 DNA polymerase to generate sticky ends. See ingredients and concentrations in table 3-8.

Table 3-8

Reaction (one well)		Stock concentration	Final concentration
[µl]			
3.3	T4 DNA Polymerase	0.3 U/µl	1 U/reaction
2	NEB buffer 2	10 ×	1×
14.7-x	H2O (sterilized)		
x (calculate with purified DNA)	100 ng vector or insert		

All the ingredients were mixed accordingly to get 20 µl reaction samples and incubated at 22 °C for 10 minutes, 10 mM dCTP (final concentration) was then added to stop the reaction. At last the cohesive vectors and inserts were obtained. With these products, the recombination reaction was performed by using RecA enzyme to envelope the final construct. The diluted RecA enzyme should be prepared freshly. All ingredients and concentrations are shown in table 3-9.

Table 3-9

Reaction (one well)	Stock concentration	Final concentration

[µl]			
2	T4 DNA Ligase buffer	10 ×	1×
1	RecA	400 ng/µl	20 ng/µl
8.5	Cohesive vector		
8.5	Cohesive insert		

All the ingredients were mixed accordingly to get 20 µl reaction samples and incubated it at 37 °C for 30 minutes at most. The samples were kept at 4 °C for the next step.

3.3.3 Transformation and culture PCR

After the SLIC, 5 µl reaction solution were transferred into 50 µl chemical competent OmniMAX^{TM}2 cells. The cells were incubated on ice for 25 minutes. Then, a heat shock for 90 s in a 42 °C water bath was applied. The cells were then placed on ice for another 2 minutes. 200 µl pre-warmed SOC medium were added to the tube and it was incubated in a shaker at 37 °C and 300 rpm. After 1 hour incubation, 50 µl cell culture were spread onto SOC agar plate with specific antibiotic and incubated overnight at 37 °C.

6-10 single colonies were picked to 20 ml LB medium containing specific antibiotic, and incubated overnight at 37 °C. On the next morning, culture PCR was performed to confirm whether the ligation is successful or not. All the ingredients are the same as normal PCR except for using cell culture as the template. Negative control and positive control were set by using water as template and construct containing accordingly sequence as template, respectively. After culture PCR is finished, an agarose gel was applied to check whether the size of the target genes is correct or not.

3.3.4 Plasmid extraction and sequencing

Based on the agarose gel, positive cell cultures were selected for plasmid extraction, and plasmids were sequenced by using the T7 forward and reverse primers, instead of the original primers designed for SLIC.

Correct plasmid was transformed into competent expression cells such as Tuner (DE3), Rosetta (DE3) pLysS and BL21 (DE3) to express accordingly proteins.

3.3.5 Protein expression and purification

A single colony from an agar plate with freshly transformed cells was picked to a 50 ml pre-culture (LB or TB medium) mixed with the required antibiotics. Cells were cultivated at 37 °C and 130 rpm for overnight. Then 5 ml overnight culture was transferred to 1000 ml fresh LB or TB medium and incubated at the same condition as before until the cell density at 600 nm (OD_{600}) reached 0.6-0.8, at which the temperature was set to 20 °C and 1 mM IPTG at final concentration was added to induce the proteins expression overnight. For full-length AmbB and AmbE proteins, 0.5 mM IPTG at 24 °C for 4 hours were used. After the induction finished, cells were harvested by centrifuging at 7,000 rpm (9,180xg) for 15 minutes at 4 °C. If not used immediately, the cells were flash frozen in liquid nitrogen and stored at -80 °C.

For purification, normally the first step applied was Ni^{2+}- immobilized metal affinity chromatography (Ni^{2+}-IMAC) (Fig 3-2), because all the proteins expressed in this project were attached with a 6×His tag at the N- or C-terminus. Cell pellets were thawed if previously frozen and then resuspended with buffer A (50 mM Tris pH 7.5, 200 mM NaCl, 10% glycerol), and lysozyme, DNase , $MgCl_2$ and protease inhibitor cocktail tablets were also added. The suspension was homogenized by stirring in the 4 °C cold room for 30 minutes. Cells were lysed by sonication for 25 minutes in an ice bath under the mode 1 second on/10 second off. After sonication, a centrifugation step was performed at 16,000 rpm (37,060xg) and 4 °C for 60 minutes to obtain the supernatant. Samples for sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) were taken from the cell lysate and supernatant.

The supernatant solution from the previous centrifugation step was loaded to a pre-packed Ni^{2+}-NTA column primed with buffer A. Then the column was washed with buffer A to elute unspecifically bound proteins. After that, the linear gradient increasing the buffer containing imidazole was applied to the column to elute the pure target protein. The eluted proteins were collected in 2 ml fractions based on the UV chromatogram and samples were analyzed by SDS-PAGE to assess the purity. If pure enough, size-exclusion chromatography was performed to obtain the final protein product used for crystallization screens. If the protein still contained impurities, then further purification steps have to be carried out, including reverse Ni^{2+}-IMAC and ion exchange chromatography.

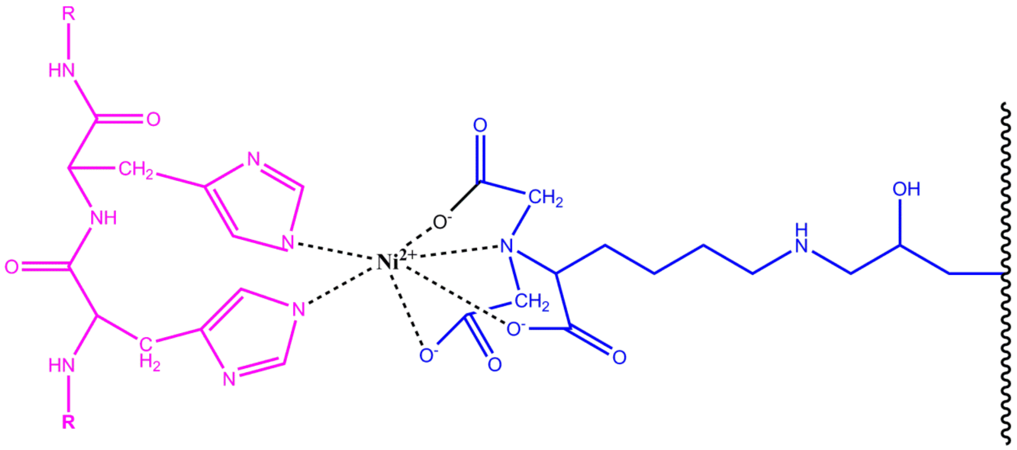


Fig 3-2: Interaction between Ni^{2+}-NTA resin and polyhistidine tag. Ni^{2+}-NTA ligand (blue) is covalently attached to a cross-linked agarose matrix for the selective purification of His-tagged proteins. The poly-His tag (magenta) can bind to Ni^{2+} ions. The bound protein is eluted by a high concentration of imidazole.

For reverse Ni^{2+}-IMAC, the His tag of proteins was removed by the respective protease. Then the protein solution was applied to a regenerated Ni^{2+}-NTA column. Generally free His tag and His tag attached protease would bind on the column and target protein without His tag would flow through. After this step, the protein should be much purer.

Ion exchange chromatography (IEX) can separate proteins of similar size that are difficult to separate by other methods, because the charge carried by proteins can be varied by changing pH of buffer (Fig 3-3). In this project the isoelectric points of all proteins are below 6, so the applied buffer pH at around 7.5 will charge protein molecule negatively. An anion exchange resin (HiTrap^{TM} 5 ml Q FF) was chosen to capture proteins under lower salt concentration. Then a linear or step gradient increasing the salt concentration of the eluting buffer was applied to elute the target proteins separately (Fig 3-3). SDS-PAGE of peak fractions was performed to assess the purity.

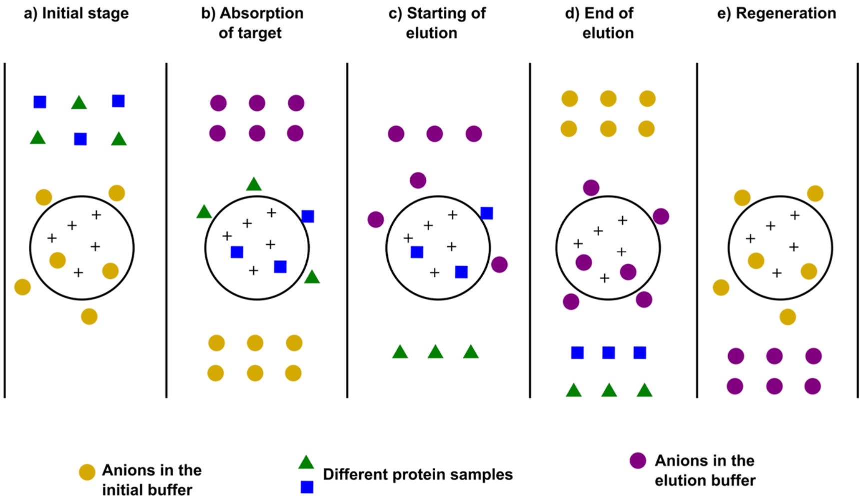


Fig 3-3: Phases of ion exchange chromatography. The positively charged resin is shown, a) initially the anions (yellow ball) from initial buffer bind to the column. b) Then two negatively charged sample ions (green triangle and blue square box) are loaded. c), d) These two samples can be separated by a gradient of ions (purple ball), based on binding strength with the resin. e) At last the resin is regenerated with the initial buffer.

A homogenous protein solution of high purity is crucial for protein crystallization. For this purpose, size-exclusion chromatography (SEC) was normally the final purification step. Different molecular weights of proteins could be separated when an aqueous solution was used to transport the samples through the column. The principle of separation by SEC is shown in Fig 3-4. Proteins were eluted isocratically in SEC buffer and collected in 2 ml fractions. Samples were analyzed by SDS-PAGE. The purest fractions were pooled and concentrated by ultrafiltration to around 15 mg/ml. Small volume aliquots were prepared and flash frozen by liquid nitrogen then stored at -80°C for further usage.

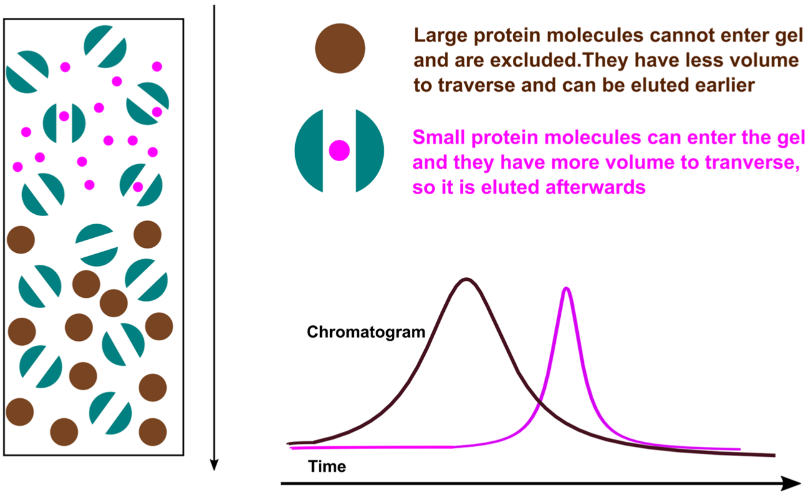


Fig 3-4: The principle of size-exclusion chromatography. The braun and magenta spheres represent proteins with different molecular weights or sizes which can be separated by travelling through the column.

The proteins studied in the course are listed in table 3-10

Table 3-10

Protein expressed in this project

Protein	Cell line	Vector	Source
His-AmbC	BL21 (DE3)	pOPINB	this project
AmbC-His	BL21 (DE3)	pET27b	Rojas Murcia, 2015
AmbD-His	BL21 (DE3)	pET27b	Rojas Murcia, 2015
His-AmbC 30-350	BL21 (DE3)	pOPINB	this project
His-AmbB	BL21 (DE3) pLysS	pET27b	Rojas Murcia, 2015
AmbE-His	BL21 (DE3) pLysS	pET27b	Rojas Murcia, 2015
His-AmbE 20-440	Tuner (DE3)	pOPINB	this project
His-AmbE 20-420	Tuner (DE3)	pOPINB	this project
His-AmbE 883-1209	Rosetta (DE3) pLysS	pOPINB	this project

3.3.6 Preparation of Seleno-*L*-methionine labeled protein

Sometimes in order to solve the phase problem of one protein structure, the anomalous diffraction dataset needs to be collected. The popular way to obtain this dataset is introducing heavy atom to the protein crystal. There are two different ways: one is soaking crystals with heavy metal atom solution, the other is purifying seleno-*L*-methionine protein and obtaining heavy atom including protein crystals. In this project, the latter method was applied.

There are several standard protocols to obtain seleno-*L*-methionine proteins, which I used is a modified one (Walden, 2010) and below are some details.

5× M9 salts stock solution was prepared by adding 64 g Na_2HPO_4-$7H_2O$, 15 g KH_2PO_4, 2.5 g NaCl and 5.0 g NH_4Cl to 800 ml MQ H_2O. After stirring, the pH was adjusted to 7.2 with NaOH and the final volume was set 1000 ml. Then the prepared solution was sterilized by autoclaving for further using.

780 ml MQ H_2O was autoclaved, then mixed with 200 ml 5× M9 salts stock solution, 20 ml 20% glucose, 2 ml 1 M $MgSO_4$ and 100 µl 1 M $CaCl_2$ to obtain around 1000 ml M9 medium.

Normal culture was prepared by adding cells from the glycerol stock to 50 ml LB medium and being cultivated at 37 °C and 130 rpm for overnight. Then 2 ml overnight normal culture was transferred to 50 ml M9 medium and incubated at the same condition as before to obtain M9 medium pre-culture. 10 ml pre-culture was further transferred to 1000 ml fresh M9 medium and incubated as before until the cell density at 600 nm (OD_{600}) reached 0.5-0.6, at which the amino acids mixture of 100 mg lysine, 100 mg phenylalanine, 100 mg threonine, 50 mg isoleucine, 50 mg leucine and 50 mg valine were supplied and incubated for extra 15 minutes. The temperature was set to 25 °C and 60 mg seleno-*L*-methionine and 1 mM IPTG at final concentration were added to induce the seleno-*L*-methionine labeled protein expression overnight.

In this study, the yield of seleno-*L*-methionine labeled protein was less than the wild type. For purification of this protein, dithiothreitol (DTT) was added into all buffers during the whole purification, which also should be finished as fast as possible.

3.3.7 Protein crystallization conditions screen and optimization

Proteins are very tiny structures measuring roughly a few nanometers, which is too small to be observed even with the strongest light microscope. X-ray diffraction of protein crystals, nuclear magnetic resonance (NMR) and cryogenic electron microscopy are main three methods to make protein structures visible. The first method was mainly applied in this project, which is also universally used to determine proteins structures, which is reflected by the fact that around 80% of all protein structures deposited in protein data bank (PDB) have been determined by crystallographic methods.

High quality protein crystals are needed to determine structures by the X-ray crystallization method. Because different proteins have different characters, there is no fixed or regular routine for protein crystallization (Rupp, 2009). To obtain crystal of a protein, plenty of conditions have to be screened and optimization of initial conditions is usually necessary. Normally, crystals will slowly grow up from saturated protein solutions, but many factors influence the crystal growth, such as pH, concentration of precipitant, salt concentration, protein concentration, temperature, additives etc. A phase diagram (Fig 3-5) is shown to explain how crystals grow.

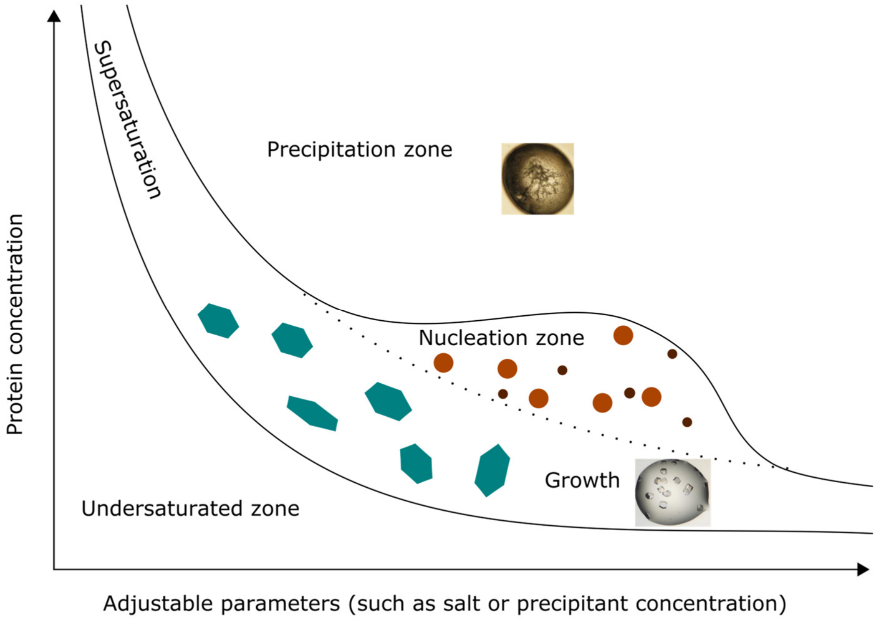


Fig 3-5: When both protein concentration and precipitant concentration are much lower, the system is in the undersaturated zone. In contrast, if they are much higher, it goes to the precipitation zone. In between, crystals can grow from the supersaturation phase. There will be plenty of tiny crystals with lower quality in the nucleation zone, while in the metastable zone, bigger and better quality crystals can emerge. The crystal stops growing if the equilibrium between unsaturated and saturated protein solution is reached (Chernov, 2003).

Vapor diffusion is the most frequently used method to achieve crystals (Alexander McPherson, 1982). The environment where protein molecules stay needs to be changed gradually to induce crystals. Two ways are normally applied to screen crystallization conditions: hanging drop and sitting drop, both are shown in Fig 3-6. The former is commonly used for high-throughput crystallization screening, and the latter one is normally applied for manual optimization of crystallization conditions. Both ways need a sealed environment in which a drop of highly concentrated protein in buffer is mixed with reservoir solution containing salts or other kinds of precipitants that decrease protein solubility, such as polyethylene glycols or alcohols. The mixing ratio is commonly 1:1, which means the concentration of salts or precipitants in the drop is lower than that in reservoir. In order to reach an equilibrium between the drop and reservoir, the water content in the drop will be reduced by vapor diffusion, resulting in gradually higher concentration of the protein inside. This process can bring proteins in the drop to supersaturated zone where potential crystals may grow.

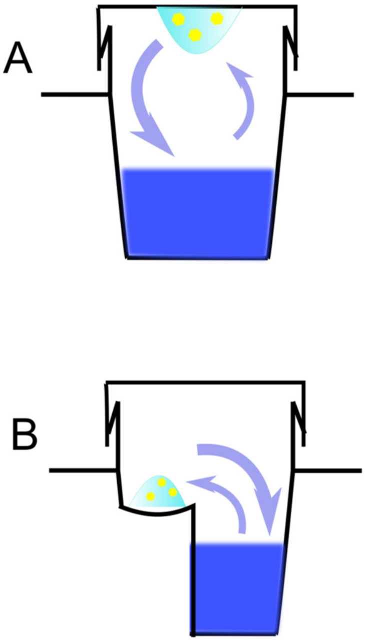


Fig 3-6: Two ways to prepare the crystallization screen. A: hanging drop. B: sitting drop.

As mentioned above, many factors can influence the crystallization procedure, which means it is not possible to predict the crystallization conditions even for a single protein. Fortunately, the most crystallization screens nowadays are performed by robots such that thousands of conditions can be screened within several hours. An automated imaging system is applied to check and track the appearance of each drop for more than one month, which can also be extended by modifying some parameters.

In most cases, the quality of crystals obtained with these initial crystallization conditions is normally insufficient, making optimization mandatory. The types and concentrations of precipitants, ratio of protein and reservoir, pH value, temperature and additive can all be varied gradient or random to obtain better diffracting crystals. Different batches of proteins can give different crystallization results. Sometimes freshly prepared protein can also make a big difference from thawed samples. If there is no crystal shown during the initial screens, significant modifications can be applied. Adding respective ligands or substrates based on the protein function can stabilize its conformation and reduce its flexibility, which will promote possibility of crystallization. Different protein constructs are also worth trying. Phyre^2 and PSIPRED can provide some clues about stable domains or core area of one protein. With these new protein sequences, crystallization screen can be performed from the very beginning.

3.3.8 Co-crystallization and crystal soaking

For protein complex structures with small ligands or substrates, co-crystallization and crystal soaking are commonly used. Before setting the crystallization screen, the protein and its substrate are mixed at a common ration 1:5, while the concentration of substrate can be increased or decreased accordingly. Optimization of crystals of complexes is the same as for the protein alone. For crystal soaking, the apo protein crystals have to be provided and transferred to a substrate-containing solution prepared with same reservoir that was used to obtain apo crystals. The soaking time and concentration of substrate have to be optimized to achieve good diffraction afterwards. The advantage of this method is that no new crystallization conditions have to be established. Soaking leads to cracking of crystals, however, since the protein changes its conformation upon binding of ligands.

If the protein is an enzyme that converts a substrate into a product, substrate analogs can be used toobtain a complex structure.

3.3.9 X-ray diffraction data collection and processing

In this project, all X-ray diffraction datasets were collected from synchrotron radiation facilities (ESRF and SLS). Most crystals were tested and screened in-house on a rotating anode X-ray generator before sending to the synchrotron. It is highly automatic nowadays to collect datasets. The collecting angel (normally 0.1° per image for 360°) and expose time and transparency can be adjusted based on the space group and quality of crystals. After collecting, the datasets were indexed and integrated by using XDS, following with pointless and aimless programs packaged in CCP4. There are some mathematical parameters that indicate the quality of datasets, such as completeness, redundancy and signal to noise ratio (I/σ). For a dataset of sufficient quality for the subsequent steps of structure determination, its completeness should be higher than 90%, R_{factor} of out shell should be lower than 50% and the signal to noise ration should be roughly above 2.

3.3.10 The principles of protein X-ray crystallography

3.3.10.1 Architecture of crystals

Crystals form by plenty of molecules highly organized in three-dimensional lattice, which can be described by crystal structure. The smallest building block which describes the entire lattice by translation operations is called *unit cell*. The geometry of the *unit cell* and symmetry characters of one crystal is described by the concept of space group. In three dimensions, there are total 230 space groups classified to 7 crystal systems, named *triclinic*, *monoclinic*, *orthorhombic*, *tetragonal*, *trigonal*, *hexagonal* and *cubic*. While as chiral molecules, proteins can only take translation and rotation as symemetric operations. Mirroring that would change chirality such as transforming *L*-amino acids into *D*-amino acids, is impossible. Accunting for this restriction, 65 of 230 space groups can be anticipated for protein crystals.

It is crucial to identify the correct space group of one protein crystal for the determination of its structure.

3.3.10.2 Diffraction of X-ray

The wavelength (λ) applied by a microscopic technique determines the resolution (d) which represents the capacity to resolve two objects as separate images. The refractive index (n), calculated by the quotient of velocity of light in vaccum and in a given material, also influences the resolution. The relationships are well explained by the below equation.

$$d = \frac{\lambda}{2 \cdot n \cdot \sin \alpha}$$

Equation: Resolution (*d*) is defined by the wavelength (*λ*) and the refractive index (*n*) and the angle of aperture of the objective (technically limited to a maximum of 90°)

In order to determine the protein structures, some intrinsic challenges have to be considered. For instance, the distance between two single atoms is roughly a few angstroms (Å), which locates totally different length level contrast to the visible light (around 390 nm – 750 nm). However, X-rays (around 0.1 Å – 100 Å) fall perfectly into the proper spectrum for determining these features.

X-ray crystallography is a technique applied for determining the molecular structure within a crystal. When X-ray photons pass the crystal, interacting with the electrons surrounding atoms of a molecule will happen. The oscillation of an electron with the same frequency as the electrical field of the photon will be induced, resulting in radiation emission from the electrons. Since these emissions happen at any random directions of the electron for all molecules in the crystal, the emitted waves will perform destructive interference resulting in zero diffraction intensity. The constructive interference leading to magnified intensity will be observed, if the path difference between these waves is equal to an integer multiple of the wavelength (see more details in section 3.3.10.3). Through rotating protein crystals in the X-ray beam, other constructive diffractions can be obtained from respective directions. All these magnified diffractions can be recorded then applied to build one three-dimensional lattice, which is inverse to that of the protein crystal, so this virtual lattice is named the *reciprocal lattice*.

3.3.10.3 Miller indices and Bragg's law

With six lattice parameters taken from the length of the cell edges (a, b, c) and the angles between them (α, β, γ), the dimensions of a unit cell can be designated (Fig 3-7A). The imaginary *Miller indices* (h, k, l) divide the unit cell into integer parallel planes. The index h represents the number of equivalent part which is cutted by this plane in the x direction. The other two indices k and l specify that in the y and z directions, respectively (Fig 3-7B).

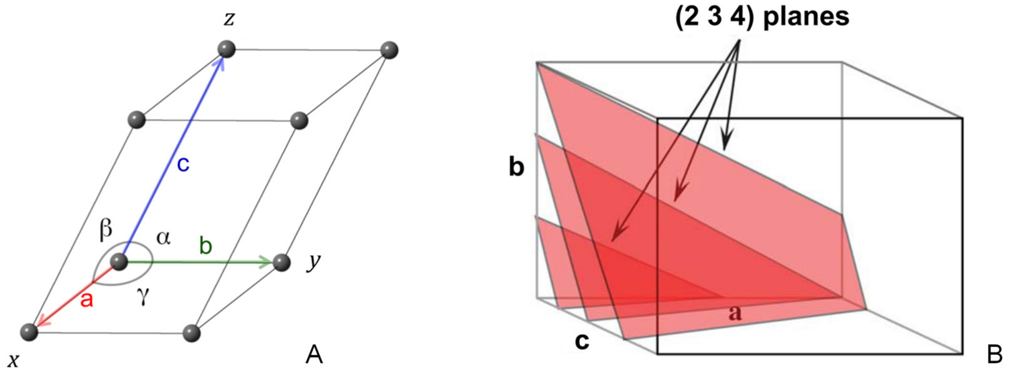


Fig 3-7. A: The typical triclinic unit cell, The parameters of edges are shown as a, b, c and the angles as α, β, γ. B: An example of reciprocal lattice planes within an unit cell, and its *Miller indices* are (2 3 4). Figure adopted from *Crystallography made crystal clear* (third edition) by *Gale Rhodes.*

These parallel and equivalent lattice planes can be taken as mirrors which reflect the incident X-rays. Under a certain incident angle θ, If the distance in path length between two beams equals an integer multiple of the wavelength λ, the constructive interference will be fulfilled, which can be well explained by *Bragg's law* (Fig 3-8).

$$2 \cdot d_{hkl} \cdot \sin\theta = n \cdot \lambda$$

Equation: Bragg's law, explaining X-ray beams under centain angles (θ) perform reflections within the crystal. λ is the variable wavelength of X-ray beam, d_{hkl} is defined by the distance between two nearby lattice planes and n means an integer.

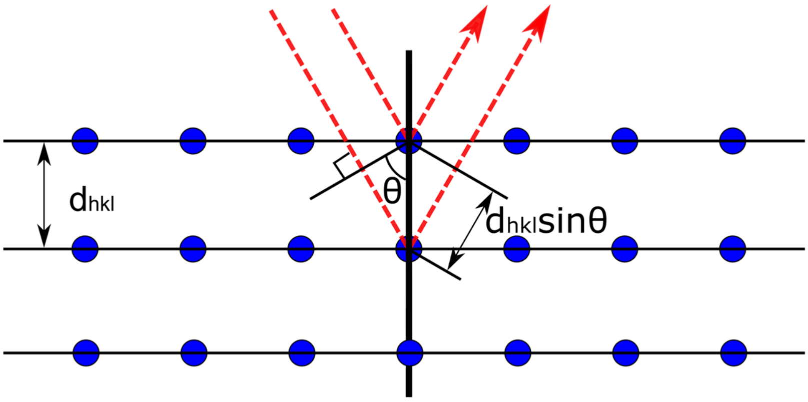


Fig 3-8. Graphical illustration of the Bragg's law. Two X-ray beams are reflected on the centain set of planes by the incident angle θ. The distance between two parallel planes is d_{hkl}, so the path difference of these two beams is calculated as $2 \cdot d_{hkl} \cdot \sin\theta$. The constructive interference only happen when it equals an integer multiple of wavelength λ.

3.3.10.4 Ewald construction

A geometric construction of reciprocal lattice points that fulfill Bragg's law is called *Ewald sphere* (Ewald, 1969). It is a three-dimensional construction with a radius $1/\lambda$, and taking the crystal as the centre. When an incident X-ray beam goes straight through the crystal, two points intersecting with the *Ewald sphere* will be identified. One is the enterance point (S_0), the other which provide reflection locates in the origin O ($0\ 0\ 0$) of the reciprocal lattice. The reflected beams that travel through the intersecting points between the reciprocal lattice points and the *Ewald sphere* will contribute constructive interferences, which can be recorded by the detector behind the crystal (Fig 3-9). As few intersecting points can be expected for each specific crystal orientation, rotation of the crystal, or its reciprocal lattice is necessary to collect enough diffractions.

The recorded diffractions coorespond to the reciprocal lattice planes, with which the original space coordinates (*h k l*) can be calculated by indexing. Then the geometry of the unit cell and the space group can also be determined.

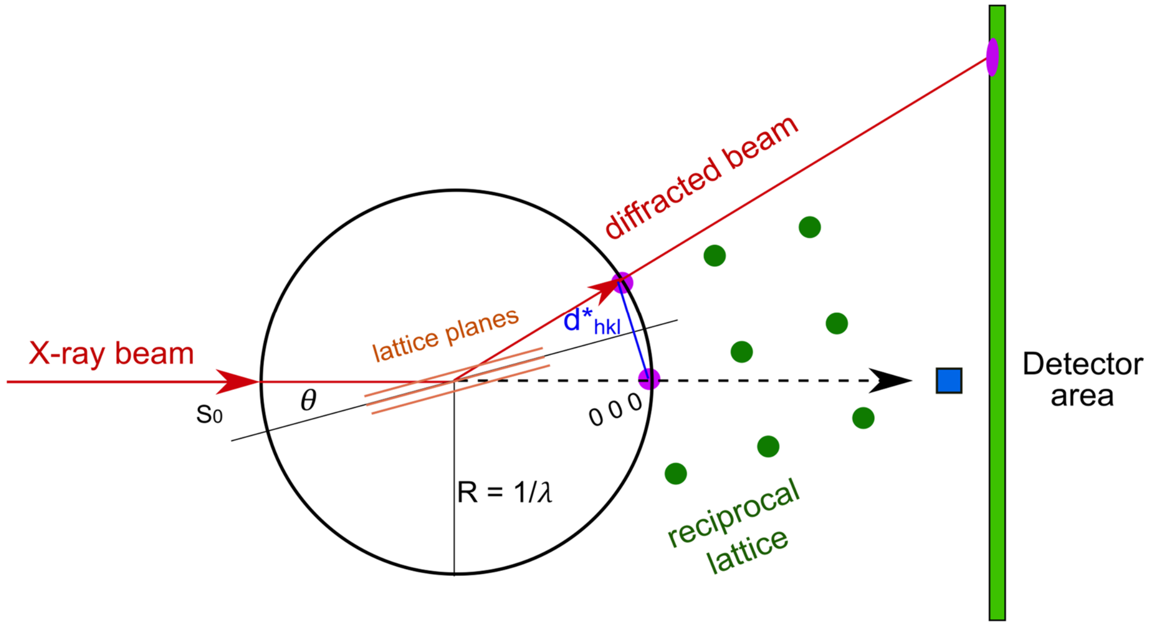


Fig 3-9. The Ewald construction. An incident X-ray beam goes through the crystal which located in the center of Ewald sphere from the entrance point S_0. It is reflected by the lattice planes, then travels through the intersecting point (0 0 1), resulting a constructive interference recorded by the detector. Rotation of crystal will produce different points into diffraction conditions. The majority of X-ray beams will go straight through the origin point (0 0 0), contributing to a strong and useless diffraction which have to be blocked by a beam stopper (blue box).

3.3.10.5 The crystallographic phase problem

For each X-ray diffraction patterns, every single atom in the unit cell makes its contribution. Every spot in the pattern is represented by one diffraction containing the magnitude and relative phase of X-ray scattered by atoms in particular direction (Fig 3-10). The contribution from all the atoms to this reflection (*h, k, l*) is called *structure factor* F_{hkl}(Cowtan, 2003).

$$F_{hkl} = \sum_{j=1}^{atoms} f_j e^{2\pi i\left(hx_j+ky_j+lz_j\right)}$$

Equation: f_j means the *atomic structure factor*.

The sum of structure fators represented by all atoms in one unit cell can be obtained through the Fourier transformation of the rediprocal lattice $F_{hkl.}$ Then the electron density for each point $\rho(x,\ y,\ z)$ in the unit cell can be calculated with the inverse Fourier transformation.

$$\rho(x,y,z) = \frac{1}{v} \sum_{hkl}^{atoms} F_{hkl} e^{-2\pi i(hx+ky+lz)}$$

Equation: v means the volume of the unit cell.

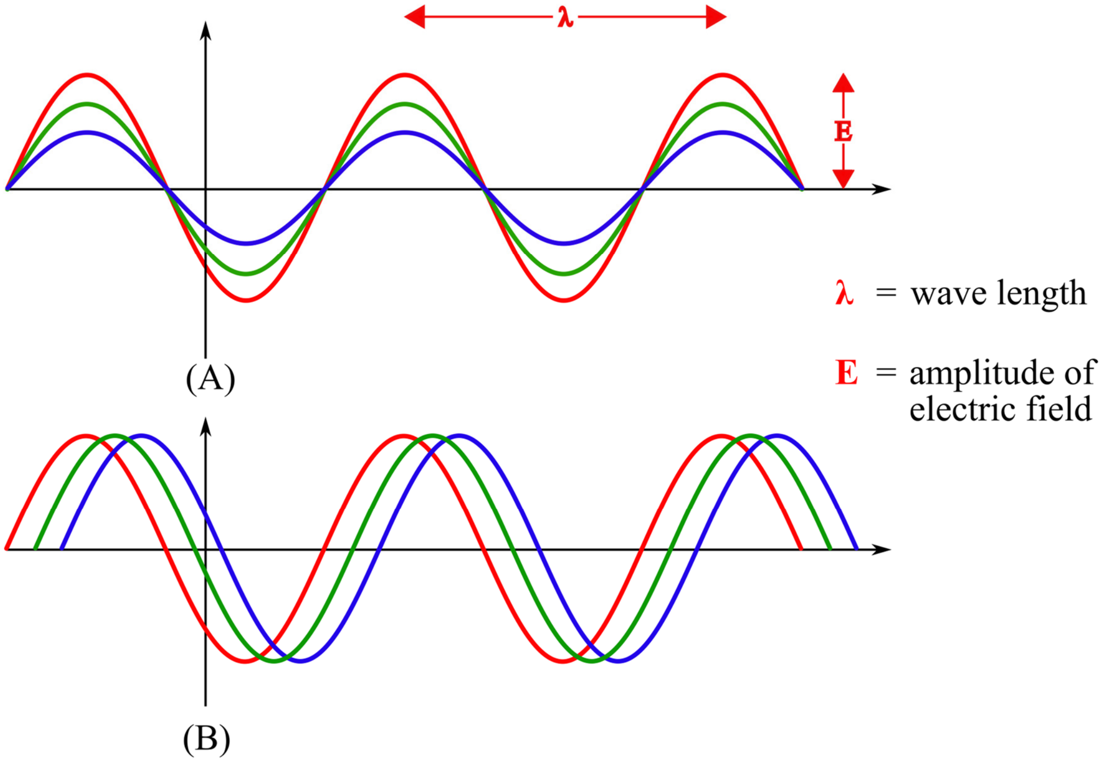


Figure 3-10. Each reflection in the diffraction pattern can be described as a wave with a certain magnitude and phase. The magnitude determines the size of the wave (A), and the phase determines where the peaks show (B).

As a wave function, each reflection recorded in the pattern contains two information, the phase $e^{i\alpha(hkl)}$ and the amplitude $/F_{hkl}/$. The detector behind the crystal will collect and record the amplitude information. However the phase information is lost during the diffraction experiments, resulting in the *phase problem* in X-ray crystallography.

The phase problem can be solved by either obtaining from a scattering experiment or borrowing from an exsiting homologous structure by molecular replacement. In this study, molecular replacement and anomalous scattering were applied to determine the structures of AmbC/D and unknown domain of AmbE, respectively, so more details about the two methods will be illustrated below.

3.3.10.6 Molecular replacement (MR)

For proteins, the primary structure determines the tertiary or quaternary structure. If one published structure shares the sequence identity with the sample protein higher than 30%, it can be taken as the phasing model to perform the molecular replacement, illustrated in Fig 3-11. If the sequence identity is not high enough, model bias will be introduced, resulting in no solution. In the process of molecular replacement, search of the model is divided into rotation and translation operations. After a few trials, a correct solution will be identified using statistical parameters such as the R factor and correlation coefficient.

$$R = \frac{\sum_{hkl} ||F_{obs}| - |F_{cala}||}{\sum_{hkl} |F_{obs}|}$$

Equation: R factor shows the difference of structure factor between the amplitude of the measured data $|Fobs|$ and that of the search model $|Fcalc|$ for every point of the translation search.

If the molecular replacement could not yield a correct solution at the beginning, it is possible to optimize the MR process. Modification of the model, such as side chain mutation can be applied. Besides, the MR pipeline programs *Balbes* (Long *et al.*, 2008) and *MrBUMP* (Keegan & Winn, 2008) are provided from CCP4 online.

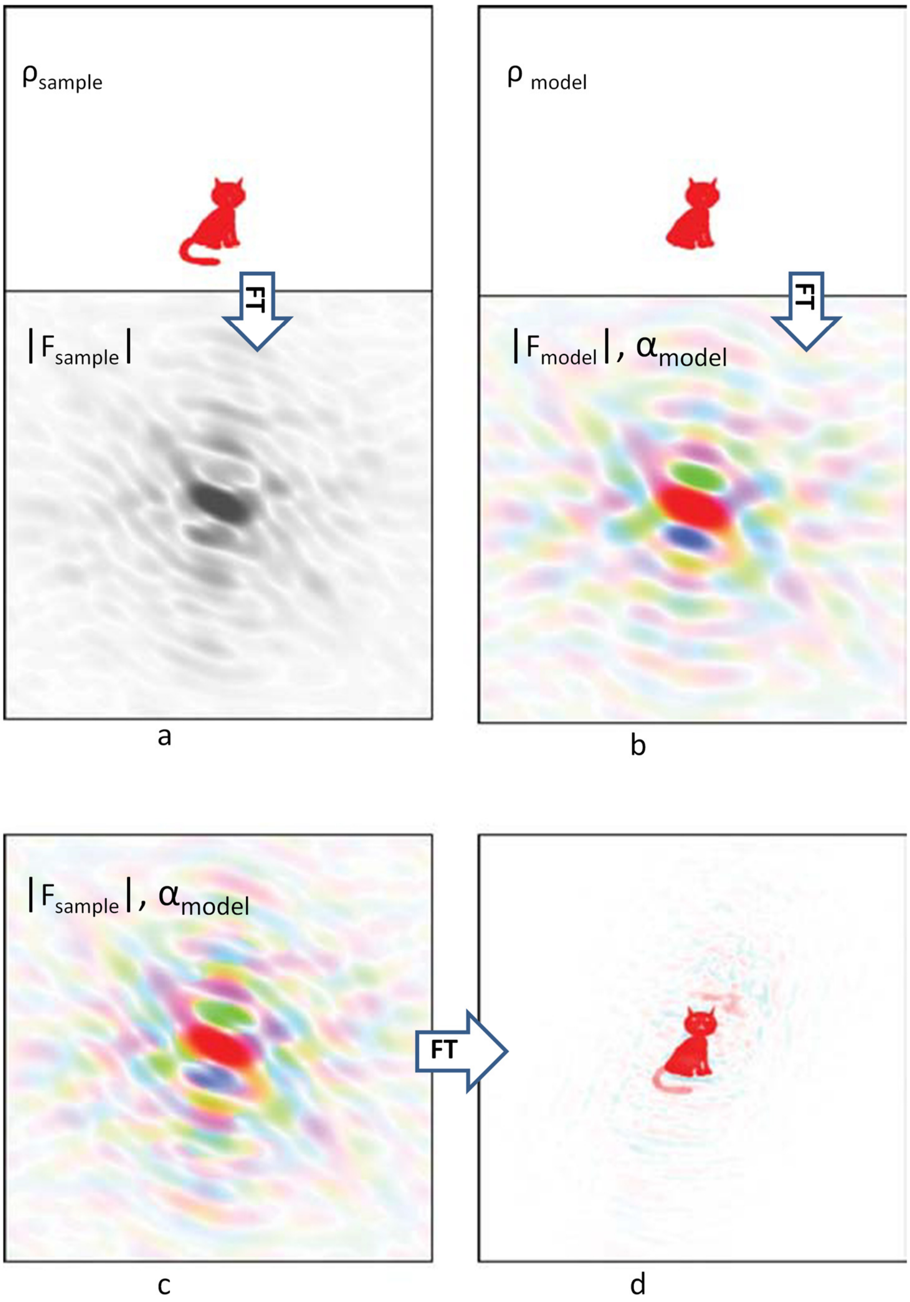


Fig 3-11. The illustration of molecular replacement. (a) Unknown structure, cat and its diffraction pattern (color is grey because the phase in absent). (b) Known structure as the searching model, cat without tail and its colorful intact information. (c) Phase (α_{model}) from the cat without tail combined with the unknown cat intensity ($|F_{sample}|$). (d) Fourier transform of (c). The tail information can be partially revealed from the intensity ($|F_{sample}|$). The original figure was adopted from *Crystallography made crystal clear* (third edition) by *Gale Rhodes.*

3.3.10.7 Anomalous diffraction

If no acceptable solution was obtained from MR, experimental phasing by introduction of anomalous scattering atoms in the structure can be accomplished. In this course, Seleno-*L*-methionine labeled protein was purified and according crystals were therefore provided to fulfill the anomalous diffraction data collection. Selenium is one of the heavier elements which exhibit absorption edges in the range of X-ray wavelengths. When certain X-ray beam goes through the selenium atom, an electron locating at a distinct shell will be activated to a higher unoccupied shell or kicked away from the atom. Absorption edge of Selenium was illustrated in Fig 3-12. Since this beam lost partial energy which was absorbed, the reflection cannot contribute to the normal scattering. As an anomalous scattering, the phase of which is not opposite in according Miller indices, resulting violation of Friedel's law.

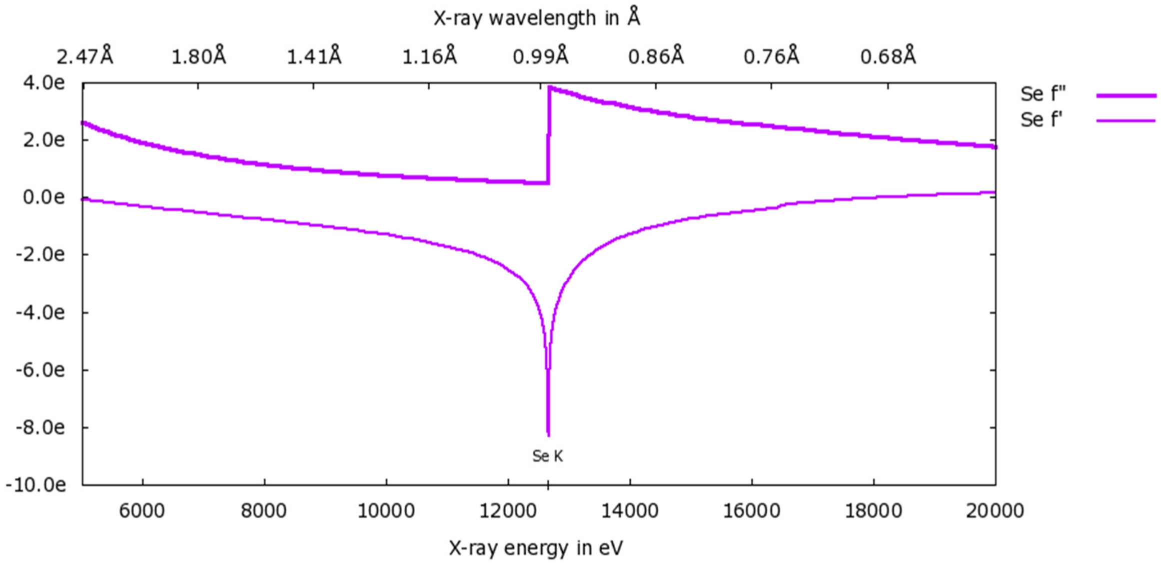


Fig 3-12: Plot of the absorption edge of selenium. When the wavelength stays between the peak (f'') and the inflection (f'), the anomalous difference can be obtained at its maximum. The plot was calculated using the subroutine library by Brennan and Cowan (S. Brennan, 1992).

$$P(\vec{u}) = \rho(\vec{r}) \cdot \rho(-\vec{r}) = \frac{1}{v}\sum\nolimits_{hkl} |\vec{F}_{hkl}|^2 e^{-2\pi i(hx+ky+lz)}$$

Equation: Patterson function. Without prior phasing information, this function can be calculated by the diffraction intensities.

Base on Patterson function, anomalous differences can be applied to determine the positions of heavier atom in the unit cell.

3.3.11 Structure refine and graphical presentations

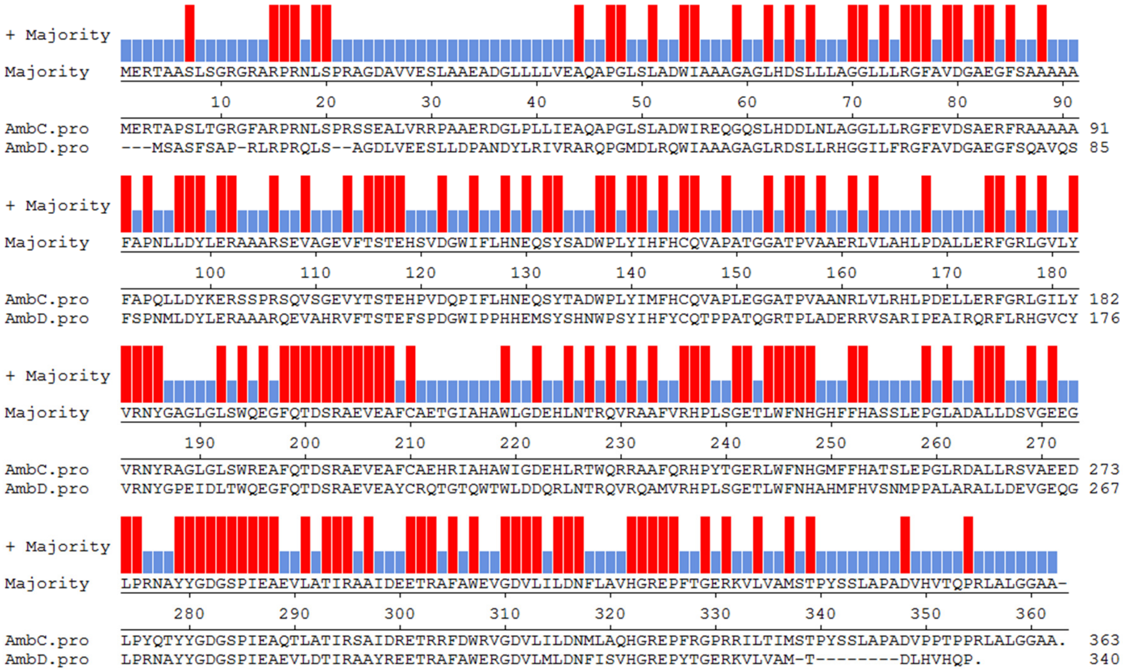
After solving the phase problem, an electron density map and an initial pdb file can be obtained, which can be applied to do refinement and manual adjustment in Coot (Emsley & Cowtan, 2004; Emsley *et al.*, 2010). MolProbity (Chen *et al.*, 2010; Davis *et al.*, 2007) was always used to validate the final structure. Analysis and graphical representation of protein structures were made by using the PyMOL software (Schrödinger, 2009).

4. Results and discussion

4.1 The structures of AmbC and AmbD

4.1.1 Characterization of AmbC and AmbD

Both AmbC and AmbD are predicted to be α-ketoglutarate-dependent dioxygenases acting as tailoring enzymes that modify the *L*-glutamate residue during AMB synthesis via a thiotemplate mechanism (Rojas Murcia *et al.*, 2015). Details about these two proteins can be found on the website http://pseudomonas.com/ (Winsor *et al.*, 2016). AmbC and AmbD both localise in the cytoplasm, and their genes overlap. The transcription direction is from *ambC* to *ambD*, and the last four base pairs of the *ambC* coding sequence, ATGA, contain the start codon (ATG) of *ambD*. AmbC contains 362 amino acids. The molecular weight is 40.7 kDa, and the isoelectric point is around 6.8. AmbD is composed of 339 amino acids, with a molecular weight of 38.7 kDa. The isoelectric point is 6.63. The sequence identity between these two proteins is nearly 49% (Fig 4-1), suggesting that their structures will share significant similarities. The PDB (Berman *et al.*, 2000) was searched for entries similar to AmbC and AmbD and two structures of an uncharacterized protein from *Arabidopsis thaliana* were identified, with PDB IDs 1Y0Z (Bitto *et al.*, 2005) and 2Q4A (Levin *et al.*, 2007). The sequences of AmbC and AmbD were also submitted to the structure prediction server Phyre2 (Kelley *et al.*, 2015), which predicted several oxygenases to be most similar (Table 4-1 and Table 4-2).



```
+ Majority
Majority    MERTAASLSGRGRARPRNLSPRAGDAVVESLAAEADGLLLLVEAQAPGLSLADWIAAAGAGLHDSLLLAGGLLLRGFAVDGAEGFSAAAAA
                     10        20        30        40        50        60        70        80        90
AmbC.pro    MERTAPSLTGRGFARPRNLSPRSSEALVRRPAAERDGLPLLIEAQAPGLSLADWIREQGQSLHDDLNLAGGLLLRGFEVDSAERFRAAAAA 91
AmbD.pro    ---MSASFSAP-RLRPRQLS--AGDLVEESLLDPANDYLRIVRARQPGMDLRQWIAAAGAGLRDSLLRHGGILFRGFAVDGAEGFSQAVQS 85

+ Majority
Majority    FAPNLLDYLERAAARSEVAGEVFTSTEHSVDGWIFLHNEQSYSADWPLYIHFHCQVAPATGGATPVAAERLVLAHLPDALLERFGRLGVLY
                  100       110       120       130       140       150       160       170       180
AmbC.pro    FAPQLLDYKERSSPRSQVSGEVYTSTEHPVDQPIFLHNEQSYTADWPLYIMFHCQVAPLEGGATPVAANRLVLRHLPDELLERFGRLGILY 182
AmbD.pro    FSPNMLDYLERAAARQEVAHRVFTSTEFSPDGWIPPHHEMSYSHNWPSYIHFYCQTPPATQGRTPLADERRVSARIPEAIRQRFLRHGVCY 176

+ Majority
Majority    VRNYGAGLGLSWQEGFQTDSRAEVEAFCAETGIAHAWLGDEHLNTRQVRAAFVRHPLSGETLWFNHGHFFHASSLEPGLADALLDSVGEEG
                 190       200       210       220       230       240       250       260       270
AmbC.pro    VRNYRAGLGLSWREAFQTDSRAEVEAFCAEHRIAHAWIGDEHLRTWQRRAAFQRHPYTGERLWFNHGMFFHATSLEPGLRDALLRSVAEED 273
AmbD.pro    VRNYGPEIDLTWQEGFQTDSRAEVEAYCRQTGTQWTWLDDQRLNTRQVRQAMVRHPLSGETLWFNHAHMFHVSNMPPALARALLDEVGEQG 267

+ Majority
Majority    LPRNAYYGDGSPIEAEVLATIRAAIDEETRAFAWEVGDVLILDNFLAVHGREPFTGERKVLVAMSTPYSSLAPADVHVTQPRLALGGAA-
                280       290       300       310       320       330       340       350       360
AmbC.pro    LPYQTYYGDGSPIEAQTLATIRSAIDRETRRFDWRVGDVLILDNMLAQHGREPFRGPRRILTIMSTPYSSLAPADVPPTPPRLALGGAA.  363
AmbD.pro    LPRNAYYGDGSPIEAEVLDTIRAAYREETRAFAWERGDVLMLDNFISVHGREPYTGERKVLVAM-T--------DLHVHQP.           340
```

Fig 4-1: Sequence alignment between AmbC and AmbD generated by Clustal W (M.A. Larkin, 2007).

Table 4-1

Top 5 templates used for structure prediction of AmbC

PDB ID	Function annotation	Sequence identity / %	Model confidence / %
1Y0Z	Clavaminate synthase-like	28	100
5HSX	putative alpha-ketoglutarate-dependent taurine dioxygenase	21	100
5J92	putative alpha-ketoglutarate-dependent taurine dioxygenase	21	100
1OTJ	Clavaminate synthase-like	16	100
5VN6	alpha-ketoglutarate-dependent taurine dioxygenase	21	100

Table 4-2

Top 5 templates used for structure prediction of AmbD

PDB ID	Function annotation	Sequence identity / %	Model confidence / %
1Y0Z	Clavaminate synthase-like	29	100
5HSX	putative alpha-ketoglutarate-dependent taurine dioxygenase	19	100
1OTJ	Clavaminate synthase-like	16	100
5J92	putative alpha-ketoglutarate-dependent taurine dioxygenase	20	100
3PVJ	alpha-ketoglutarate-dependent taurine dioxygenase	16	100

4.1.2 Production and purification of AmbC and AmbD

The genes AmbC and AmbD from *Pseudomonas aeruginosa* were cloned into the vector pOPINB. The proteins were fused to an N-terminal His tag and a PreScission protease cutting site. Both proteins were expressed and purified as usual. However, AmbD produced with the pOPINB construct was insoluble. The Ni^{2+}-IMAC and final size-exclusion chromatography (SEC) purifications of AmbC were identified by SDS-PAGE (Fig 4-2). The pooled eluate fractions were concentrated to a final concentration of 22 mg/ml. Small volume aliquots were prepared and flash frozen and stored at -80 °C.

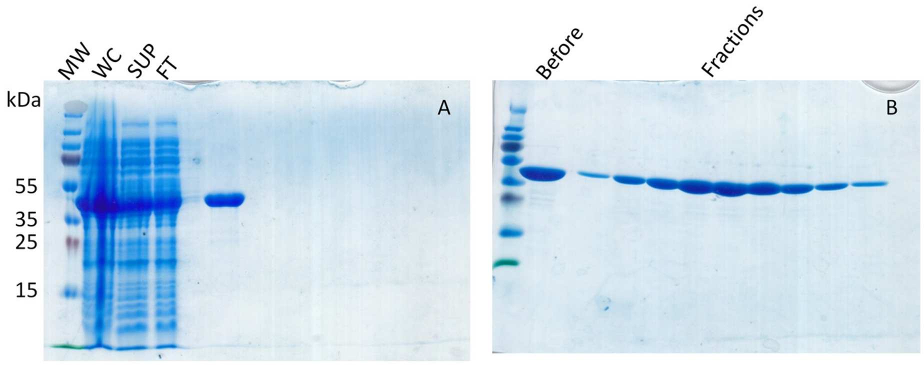


Fig 4-2: SDS-PAGE followed by InstantBlue staining for AmbC. (A): Ni^{2+}-IMAC purification. MW: Molecular weight marker. WC: whole cell lysate. SUP: supernatant of

lysate upon centrifugation. FT: flowthrough. (B): Size-exclusion chromatography (SEC) of AmbC. Before: the sample before SEC. Fractions: peak fractions.

Recombinant full-length *ambC* and *ambD* constructs with C-terminal His tag in the vector pET27b were obtained from Prof. Cornelia Reimman. The constructs were named AmbC-His and AmbD-His, respectively. The same expression and purification protocol as for *ambC* in pOPINB before was applied. SDS-PAGE analysis is shown in Fig 4-3. Peak fractions were pooled and concentrated to 12 mg/ml for AmbC-His and 8 mg/ml for AmbD-His, repectively. Small volume aliquots were prepared and flash frozen and stored at -80 °C.

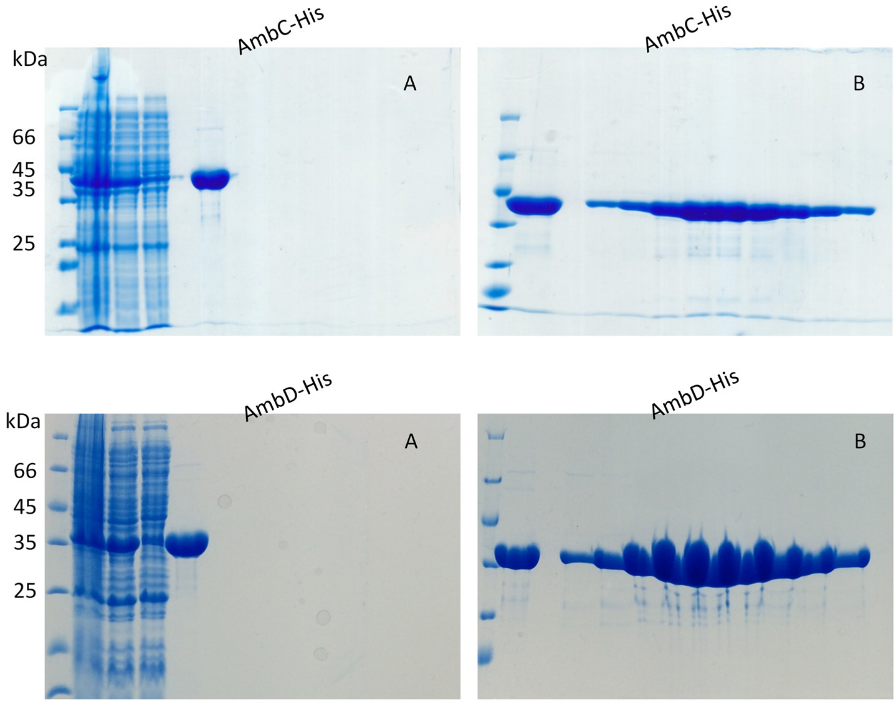


Fig. 4-3: SDS-PAGE followed by InstantBlue staining for purification of AmbC-His and AmbD-His. (A): Ni^{2+}-IMAC purification of AmbC-His and AmbD-His, respectively. (B): Size-exclusion chromatography (SEC) of AmbC-His and AmbD-His, repectively. Annotations are the same as Fig 4-2.

4.1.3 Crystallization of AmbD-His and X-ray diffraction experiments

Initial crystallization screens were set up for AmbC, AmbC-His and AmbD-His. One hit was identified for AmbD-His (JCSG II E7) (Fig. 4-4). The shape and surface of the crystals were not perfect. However, reproduction and optimization were unsuccessful and no crystals were found even after excessive trials. Therefore three initial crystals were sent to beamline P11 in DESY (Burkhardt, 2016), allowing collection of two datasets. 3600 images of 0.1° rotation were collected on a Pilatus 6M detector at a wavelength of λ = 1.0332 Å. The better dataset had been analyzed and detailed statistics are shown in Table 4-3.

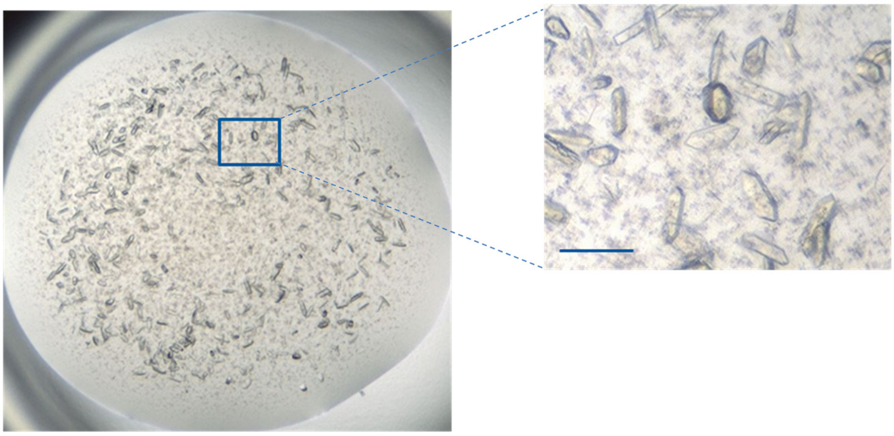


Fig. 4-4: Crystals of AmbD-His obtained in the initial screen (0.1 M imidazole pH 8.0, 0.2 M zinc acetate and 40% PEG600). Scale bar is 100 µm.

Table 4-3: Data collection statistics for AmbD-His generated with aimless program (Evans, 2011; Evans PR, 2013). Values in parentheses are for the highest resolution shell.

Dataset	AmbD-His
Wavelength (Å) / beamline	1.0332 / DESY, P11
Space group	$P2_12_12_1$
Resolution range (Å)	44.64-2.9 (3.08-2.9)
Unit cell parameters (°)	90, 90, 90
(Å)	76.62, 86.04, 104.44
Mosaicity (°)*	0.27
Rpim	8 (49.3)
Total number of observations	412152 (68674)
Total number unique	15896 (2520)
Mean I / σ (I)	8.7 (1.8)
Mn (I) half-set correlation CC (1/2) (%)	99.5 (88.7)
Completeness (%)	100 (100)
Multiplicity	25.9 (27.3)

*Mosaicity value is from XDS (Kabsch, 2010)

The crystal of AmbD-His diffracted to a resolution of 2.9 Å at P11 beamline in DESY and reflections were indexed in space group $P2_12_12_1$ with cell dimensions a = 76.62, b = 86.04, c = 104.44. There were two molecules in one asymmetric unit according to Matthews_coef calculation in CCP4 (Kantardjieff & Rupp, 2003; Matthews, 1968). No solution could be found by molecular replacement methods with this dataset. Therefore in order to obtain initial phases, selono-*L*-methionine-labelled AmbD-His was purified and crystallization screens were set up, but crystallization was not achieved. However, the structure of AmbD-His was eventually solved by molecular replacement with AmbC 30-350 (chapter 4.1.10).

4.1.4 Construction and purification of AmbC 30-350

Since full-length AmbC 1-362 could not be crystallized, further analysis of its characteristics was performed. Sequence analysis results from PSIPRED and the FoldIndex online server (Buchan *et al.*, 2013; Jaime Prilusky, 2005) are shown in Fig 4-5 and Fig 4-6, respectively. The N-terminal 30 amino acids and C-terminal 10 amino acids indicate a high possibility of disorder, which could decrease the tendency to form crystals. As a consequence, a truncated version of AmbC (AmbC 30-350) has been prepared.

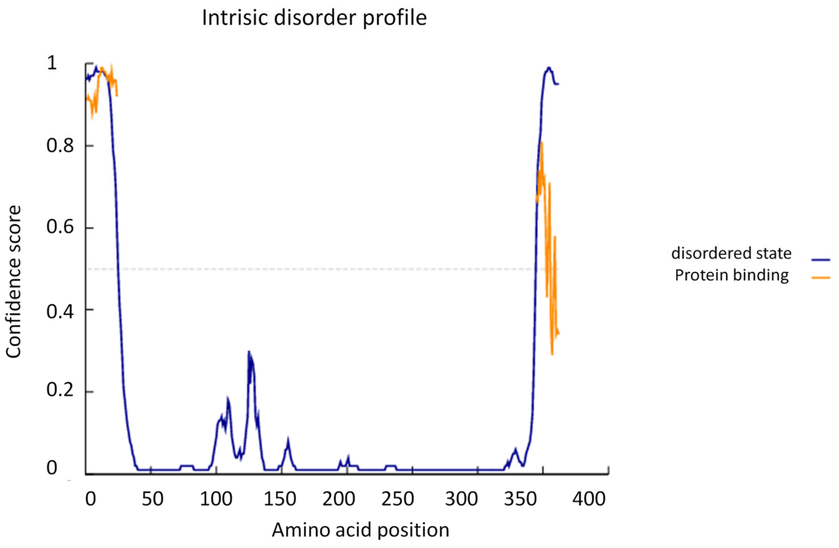


Fig 4-5: DISOPRED analysis of AmbC. The blue line represents the likelihood of disorder. Amino acids located at the termini are predicted as disordered. The orange line shows the confidence of disordered protein binding residue predictions.

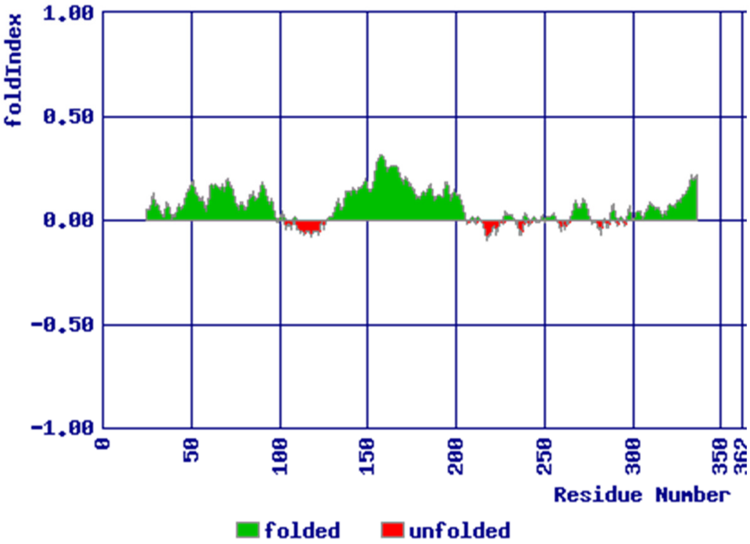


Fig 4-6: FoldIndex of AmbC. The missing green signal for the N-terminal 30 amino acids and the C-terminal 10 amino acids indicate that these regions are flexible. The red unfolded sequences represent predicted flexible loops.

The coding sequence of residues 30-350 of AmbC was PCR amplified by using the pET27b *ambC* plasmid as the template. The PCR product was inserted between the KpnI and HindIII restriction sites of pOPINB. The protein AmbC 30-350 was expressed and purified following the same routine that was used for the full-length AmbC. The Ni^{2+}-IMAC and SEC purification results were identified by SDS-PAGE (Fig. 4-7)

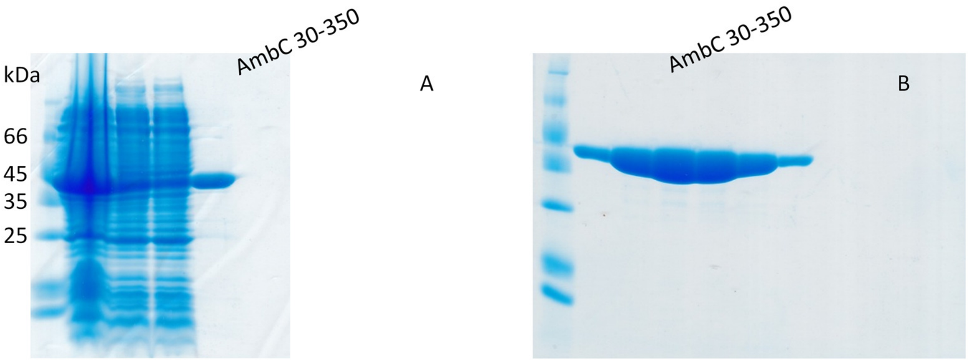


Fig 4-7: SDS-PAGE followed by InstantBlue staining for purification of AmbC 30-350. (A): Ni^{2+}-IMAC purification of AmbC 30-350. (B): Size-exclusion chromatography of AmbC 30-350. Annotations are the same as Fig 4-2.

4.1.5 Crystallization of AmbC 30-350

Two initial hits with similar precipitants were identified for AmbC 30-350 (PEG II Suite D8 and PEG II Suite D11). One condition contained 0.1 M HEPES pH 7.5, 0.2 M $CaCl_2$ and 30% PEG4000, and the other 0.1 M Tris-HCl pH 8.5, 0.2 M $MgCl_2$ and 30% PEG4000. Random optimization screens were designed and larger crystals were obtained in 0.1 M Tris-HCl pH 8.0, 0.25 M $MgCl_2$ and 30% PEG4000 (Fig 4-8).

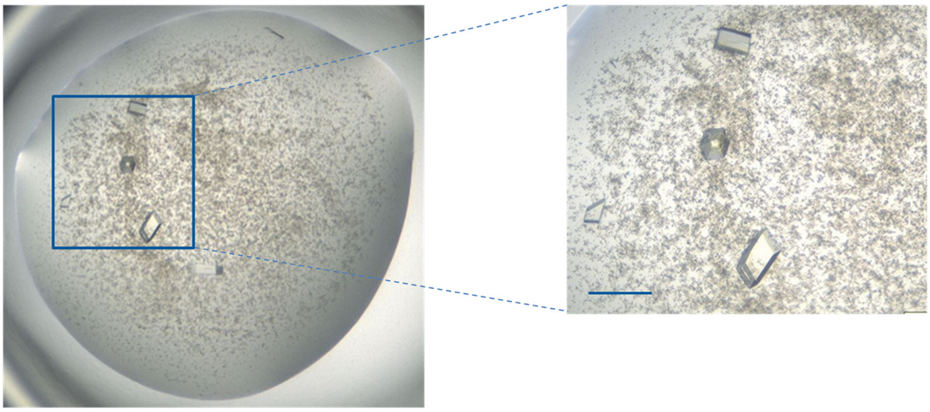


Fig. 4-8: Crystals of AmbC 30-350 obtained from optimization condition (0.1 M Tris-HCl pH 8.0, 0.25 M MgCl2 and 30% PEG4000). Scale bar is 100 µm.

Given that the crystallization condition contains 30% PEG4000, there was no need for additional cryoprotection. Three crystals were flash frozen in liquid nitrogen and sent to beamline ID30B in ESRF (Mueller-Dieckmann, 2015), where three datasets were collected. 3600 images of 0.1° rotation were collected on a Pilatus 6M detector at a wavelength of λ = 0.976252 Å. Detailed statistics of one of the best dataset are shown in Table 4-4.

Table 4-4: Data collection statistics for AmbC 30-350 calculated by the aimless program (Evans, 2011; Evans PR, 2013). Values in parentheses are for the highest resolution shell.

Dataset	AmbC 30-350
Wavelength (Å) / beamline	0.976252 / ESRF, ID30B
Space group	$P2_12_12_1$

Resolution range (Å)	108.99-1.77 (1.87-1.77)
Unit cell parameters (°)	90, 90, 90
(Å)	70.67, 76.98, 108.99
Mosaicity (°)*	0.195
Rpim	4 (43)
Total number of observations	732764 (106391)
Total number unique	58479 (8430)
Mean I / σ (I)	12.2 (2.3)
Mn (I) half-set correlation CC (1/2) (%)	99.8 (78.5)
Completeness (%)	99.9 (99.8)
Multiplicity	12.5 (12.6)

*Mosaicity value is from XDS (Kabsch, 2010)

The crystal of AmbC 30-350 diffracted to a resolution of 1.77 Å and reflections were indexed in space group $P2_12_12_1$ with cell dimensions a = 70.67, b = 76.98, c = 108.99. There were two molecules in the asymmetric unit with 40.1% solvent according to Matthews_coef calculation in CCP4 (Kantardjieff & Rupp, 2003; Matthews, 1968)

4.1.6 Structure determination and refinement of AmbC 30-350

The initial structure of AmbC 30-350 was determined by molecular replacement using the structure of a putative Fe (II)/α-ketoglutarate-dependent enzyme (PDB ID: 1Y0Z) from *Arabidopsis thaliana* (Bitto *et al.*, 2005) as the search model. Its sequence identity to AmbC 30-350 is around 28% (see also Table 4-1). The Phaser program (Bunkoczi *et al.*, 2013; McCoy *et al.*, 2007) found one solution, with which AutoBuild (Terwilliger *et al.*, 2008) was used to obtain a better structure. Refinement with phenix.refine (Headd *et al.*, 2014; Terwilliger, 2003) and manual adjustments performed in Coot (Emsley & Cowtan, 2004) followed to polish the final structure of AmbC 30-350 with R_{work} = 0.1847 and R_{free} = 0.2171 with 98.42% of the residues in Ramachandran-favored regions and a *MolProbity* (Chen *et al.*, 2010; Davis *et al.*, 2007) score of 1.10 (Table 4-5)

Table 4-5: Refinement statistics of AmbC 30-350 generated with Phenix (Diederichs & Karplus, 2013). Values in parentheses are for the highest resolution shell.

Dataset	AmbC 30-350
Resolution range (Å)	59.3 - 1.772 (1.835 - 1.772)
Total reflections	749599 (73838)
Unique reflections	58394 (5758)
Reflections used in refinement	58366 (5758)
Reflections used for R-free	2853 (287)
R_{work}	0.1847 (0.2845)
R_{free}	0.2171 (0.3135)
Number of non-hydrogen atoms	5034
macromolecules	4623
ligands	8
solvent	403
Protein residues	576
RMS(bonds)	0.004
RMS(angles)	1.04
Ramachandran favored (%)	98.42
Ramachandran outliers (%)	0
Rotamer outliers (%)	0.21
Clashscore	3.07
Average B-factor	34.67
macromolecules	34.48
ligands	35.82
solvent	36.86
Number of TLS groups	7
MolProbity score*	1.10

*MolProbity value is from MolProbity (http://molprobity.biochem.duke.edu/) (Chen *et al.*, 2010; Davis *et al.*, 2007)

4.1.7 Overall structure of AmbC 30-350

The final refined model of AmbC 30-350 contains two chains (A & B) in the asymmetric unit. Both of them could not be traced. Several amino acids at the N- and C-terminus are missing, and one gap exists in between each chain. For both chains, all the residues that could be modeled are shown in table 4-6.

Table 4-6: Residues of AmbC 30-350 that can be modeled at the resolution of 1.77 Å. For clarity, the first residue of the truncated AmbC 30-350 is numbered 1.

Chains	Start	End	Gaps
A	G8	L315	D69-H90
B	G8	P317	D69-H90

According to size-exclusion chromatography, AmbC 30-350 was expected to be a monomer in solution. The presence of two monomers (Fig 4-9A) in the asymmetric unit is the consequence of crystal packing. Superposition of the two molecules resulted in an almost complete overlap with an r.m.s.d. value of 0.716 Å (Fig 4-9B). Chain B would be shown for the following analysis.

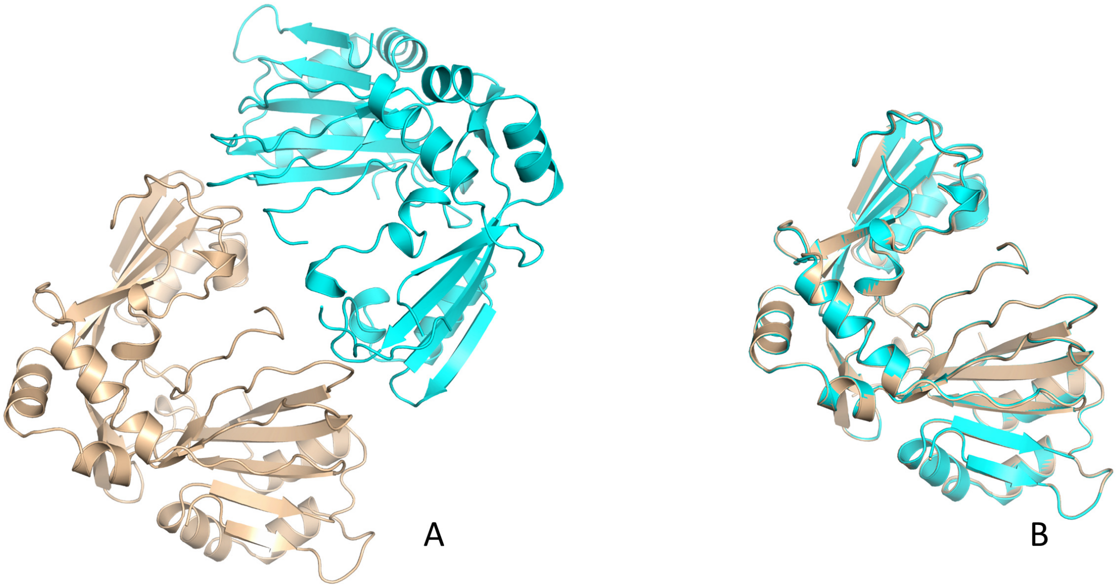


Fig 4-9: Overall structure of AmbC 30-350. A: Two chains present in one asymmetric unit. Chain A is colored in wheat and chain B in cyan. B: superposition of the two chains.

4.1.8 Secondary and tertiary structure of AmbC 30-350

The structure of AmbC 30-350 was submitted to PDBsum (Laskowski *et al.*, 2018), which generated the information about secondary structure and topology shown in Fig

4-10 and Fig 4-11. The stereo view of overall structure was prepared with PyMOL (Fig 4-12). Each AmbC 30-350 monomer contains 5 β-sheets (17 strands) and 12 α-helixes. As predicted as Fe/α-ketoglutarate-dependent enzymes, the core of AmbC 30-350 consists of a jelly roll β-sandwich of anti-parallel sheets. Nearby the core area, conserved residues His^{99} and Glu^{101} located on one β-strand and extended loop respectively, and the third conserved residue coordinating metal ion binding was His^{293}, which located on its anti-parallel β strand.

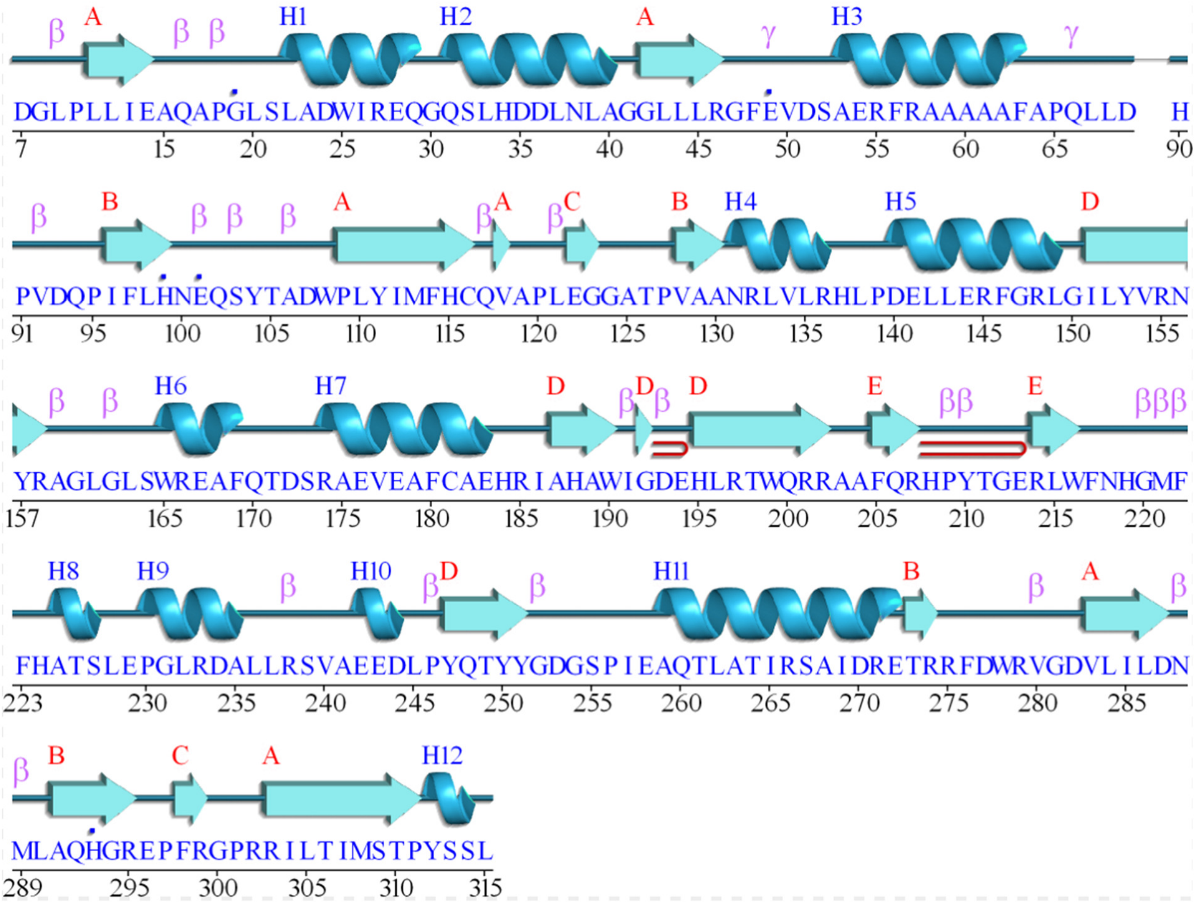


Fig 4-10: Primary and corresponding secondary structure of AmbC 30-350. The 12 α-helixes are shown as blue letters H1-H12, and 5 β-sheets formed by continuous β-strands are shown as red letters A-E. β means beta turn, γ means gamma turn, red clip means beta hairpin, blue dot means residue contacts to metal or ligand.

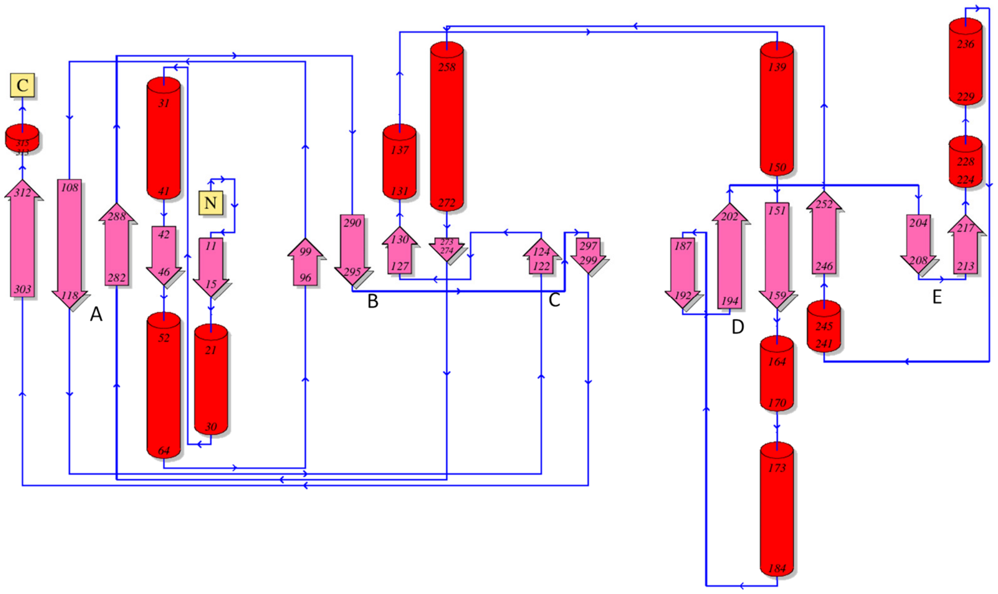


Fig 4-11: Topology of AmbC 30-350. Red columns represent 12 α-helixes, and pink arrows represent 17 β-strands classified to 5 β-sheets (A, B, C, D and E). The conserved metal binding residues His^{99}, Glu^{101} and His^{293} are located near the B β-sheet.

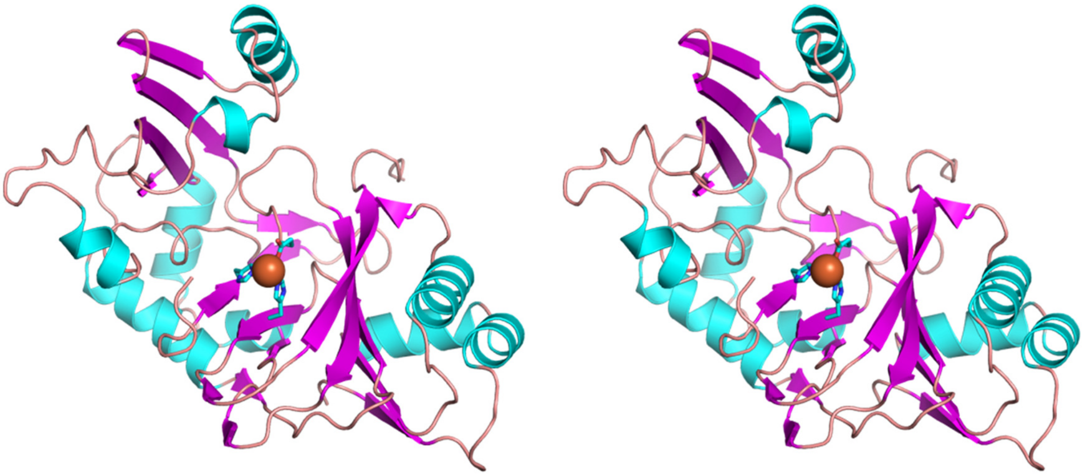


Fig 4-12: Stereo view of an AmbC 30-350 monomer shown as cartoon. The helices are colored in cyan, the β-strands in magenta and the loops in salmon. Metal is shown as spheres colored in orange, and conserved residues His^{99}, Glu^{101} and His^{293} are shown as sticks colored by elements (carbon atoms were colored in cyan).

4.1.9 The active site of AmbC 30-350

The structure of AmbC 30-350 was submitted to the metal binding site validation server CheckMyMetal (CMM) (Zheng *et al.*, 2014; Zheng *et al.*, 2017). CMM applies up to eight parameters to evaluate the inherent consistency and geometrical arrangement of each metal-binding site. The threshold values for all parameters are determined by using a benchmark data set. The values of parameters are shown and labeled as “outlier”, “borderline” or “acceptable” by coloring the table cells as red, amber or green respectively.

At the beginning, the structure was submitted with iron as the element in the metal binding site. The iron-nitrogen distances obtained with CMM were quite consistent with data from the Cambridge Structural Database (CSD), but the iron-oxygen distances were unusual (Fig 4-13). Some results generated in CMM report did not support the presence of iron in the structure (Table 4-7). Nickel was then tested by CMM instead of iron. The resulting nickel-nitrogen and nickel-oxygen distances were much closer to the CSD data than the iron distance (Fig 4-14). CMM report was fully supporting the presence of nickel in the structure (Table 4-8). It is therefore more likely that nickel rather than iron is the metal present in the structure of AmbC 30-350. It is likely that AmbC 30-350 captured nickel during the purification by Ni^{2+}-IMAC.

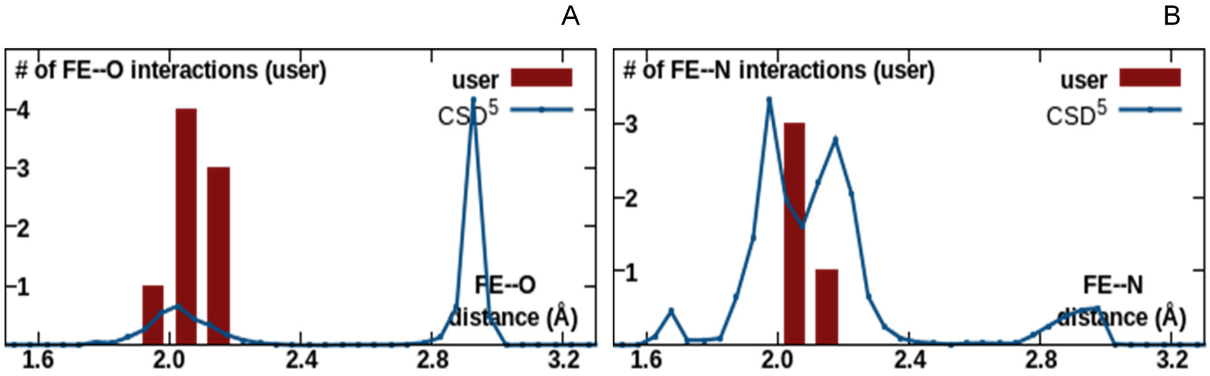


Fig 4-13: Iron metal-ligand distance distribution in the AmbC 30-350 structure compared with CSD (Groom *et al.*, 2016). A: Iron-oxygen (eight in total) distances. B: Iron-nitrogen (four in total) distances. The blue curve represents the distance distribution from CSD, a database of very high-resolution X-ray crystallography structures of small molecules. User means the submitted structure.

Table 4-7: Selected parameters from the CMM report of AmbC 30-350 with iron

ID	Metal	B factor (env.)	Ligands	valence	Geometry	gRMSD (°)	Vancancy

E: 1	FE	22 (23.6)	O_4N_2	2.5	*Octahedral*	3.7°	0
E: 2	FE	*22.3 (26.9)*	O_4N_2	2.5	*Octahedral*	4.9°	0

Ligand: Not applicable Outlier *Borderline* Acceptable

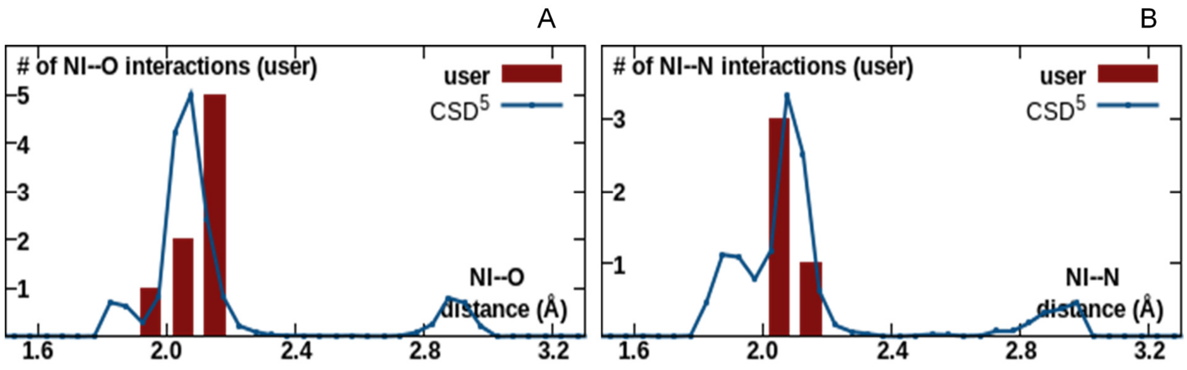


Fig 4-14: Nickel metal-ligand distance distribution in the AmbC 30-350 structure compared with CSD (Groom *et al.*, 2016). A: Nickel-oxygen (eight in total) distances. B: Nickel-nitrogen (four in total) distances. The blue curve represents the distance distribution from CSD, a database of very high-resolution X-ray crystallography structures of small molecules. User means the submitted structure.

Table 4-8: chosen parameters generated from CMM report of AmbC 30-350 with nickel

ID	Metal	B factor (env.)	Ligands	valence	Geometry	gRMSD (°)	Vancancy
F: 1	Ni	21.7 (24.1)	O_4N_2	2.2	Octahedral	5.1°	0
F: 2	Ni	24.6 (26.9)	O_4N_2	2.1	Octahedral	5.1°	0

Ligand: Not applicable Outlier *Borderline* Acceptable

Nickel was then added with Coot and refinement was performed with Phenix, finally it coordinated perfectly with the residues His^{99}, Glu^{101} and His^{293}. Because of the absence of the co-substrate α-ketoglutarate, the three left coordinating sites of metal were all occupied by water molecules. All coordination ligands were arranged in a slightly distorted octahedral configuration (Fig 4-15, Fig 4-16 and Fig 4-17). The distances between the nickel atom and coordinated atoms are around 2.0 Å.

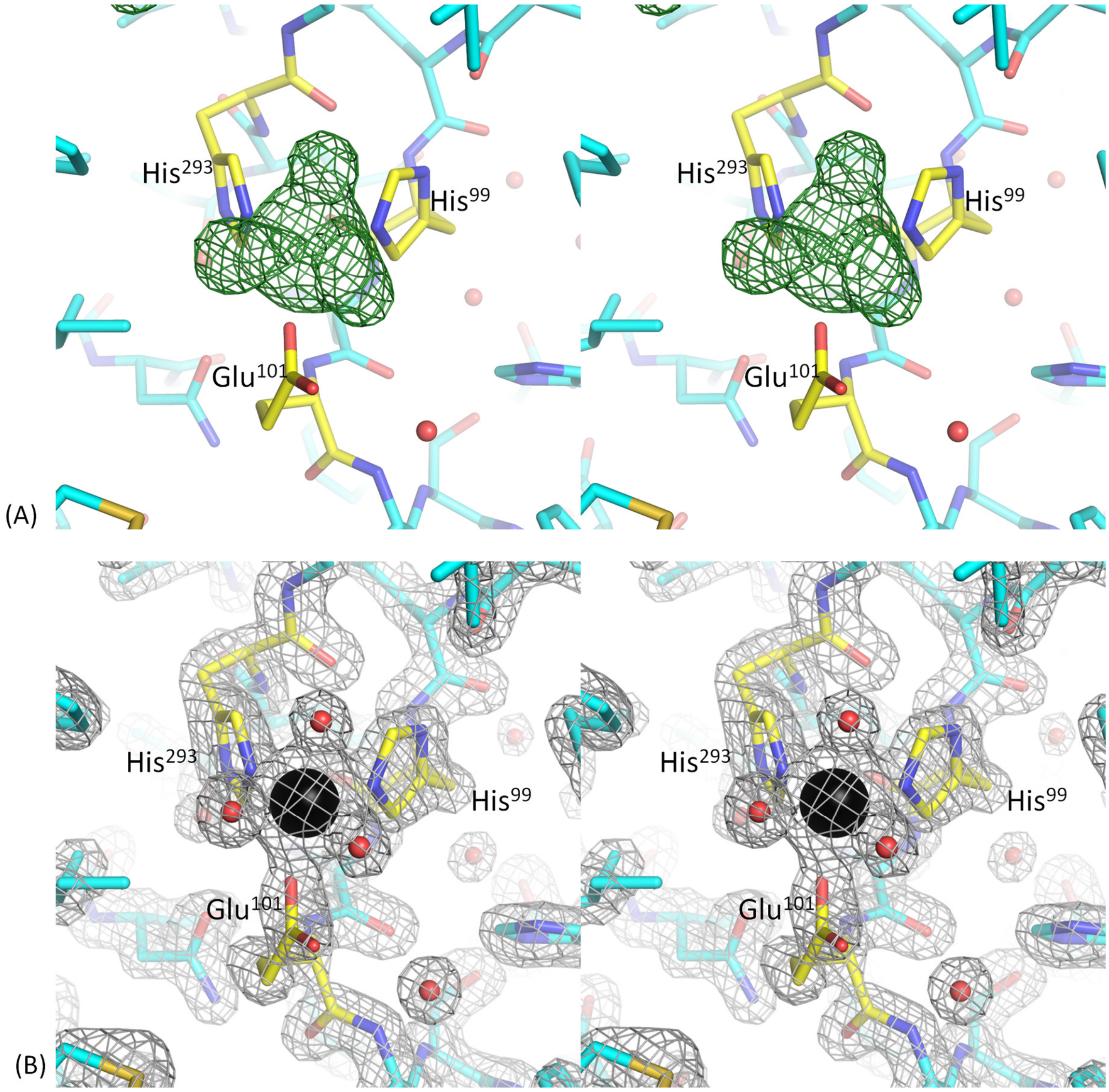


Fig 4-15: Stereo view of the active site of AmbC 30-350 and electron density maps. (A): $F_{obs} - F_{calc}$ difference electron density map (green mesh) contoured level at 4.0 σ, without the metal nickel and three water molecules. (B): The $2F_{obs} - F_{calc}$ map (grey mesh) calculated with metal nickel (black sphere) and three water molecules contoured level at 2.0 σ. The water molecules are shown as spheres colored in red. Three coordinated residues His^{99}, Glu^{101} and His^{293} are labeled and colored by elements (carbon atoms are colored in yellow).

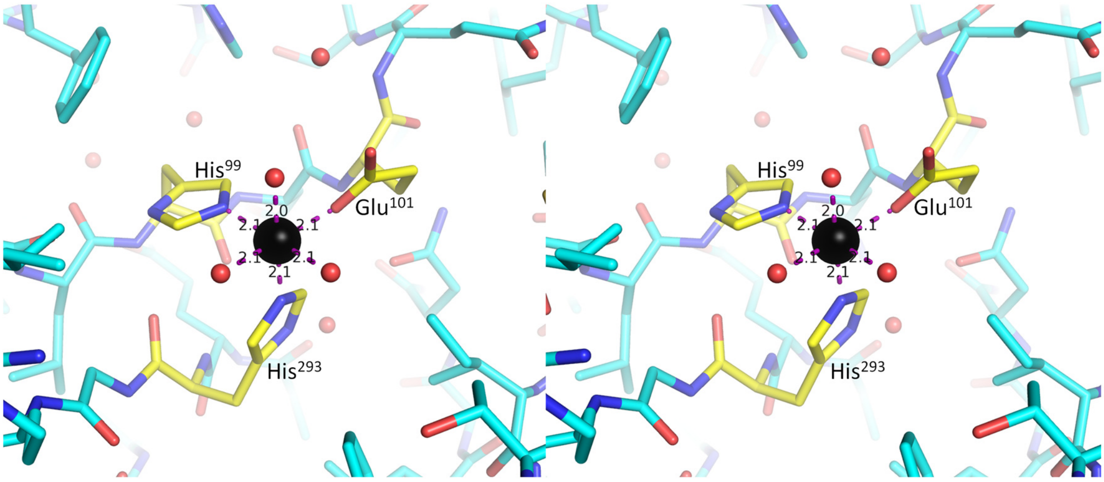


Fig 4-16: Stereo view of the active site of AmbC 30-350 built with a nickel ion. The distances between the nickel atom and all coordinated ligands are indicated in Å and are measured with PyMOL. The three coordinated residues are labeled and colored by elements (carbon atoms are colored in yellow)

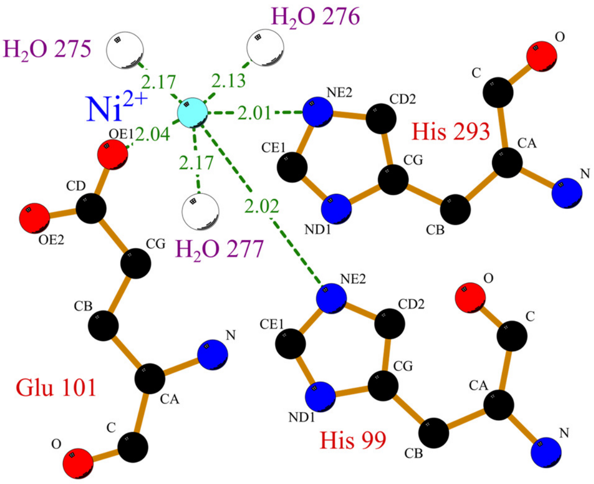


Fig 4-17: Arrangement of the active site of AmbC 30-350. The central nickel atom is coordinated by three residues and three water molecules. The figure was generated from PDBsum (Laskowski *et al.*, 2018).

4.1.10 Structure determination and refinement of AmbD

Taking the structure of AmbC 30-350 as a model, the structure of AmbD could be determined by molecular replacement. Phase-MR in Phenix was used to find a high quality solution. AutoBuild was run to obtain a decent model. Refinement with phenix.refine (Headd *et al.*, 2014; Terwilliger, 2003) and manual adjustments performed in Coot (Emsley & Cowtan, 2004) were followed to polish the structure of AmbD. This resulted in R_{work} = 0.2047 and R_{free} = 0.2591 with 97.45% of the residues in Ramachandran-favored regions and a *MolProbity* (Chen *et al.*, 2010; Davis *et al.*, 2007) score of 1.44 (Table 4-9).

Table 4-9: Refinement statistics of AmbD generated with Phenix (Diederichs & Karplus, 2013). Values in parentheses are for the highest resolution shell.

Dataset	AmbD
Resolution range (Å)	44.64 - 2.9 (3.004 - 2.9)
Total reflections	419562 (42910)
Unique reflections	15855 (1555)
Reflections used in refinement	15801 (1547)
Reflections used for R-free	296 (34)
R_{work}	0.2047 (0.3143)
R_{free}	0.2591 (0.3835)
Number of non-hydrogen atoms	5115
macromolecules	5110
solvent	5
Protein residues	633
RMS(bonds)	0.005
RMS(angles)	1.09
Ramachandran favored (%)	97.45
Ramachandran outliers (%)	0
Rotamer outliers (%)	0.57
Clashscore	6.02
Average B-factor	56.6
macromolecules	56.62
solvent	41.1
MolProbity score*	1.44

*MolProbity value is from MolProbity (http://molprobity.biochem.duke.edu/) (Chen *et al.*, 2010; Davis *et al.*, 2007)

4.1.11 Overall structure of AmbD

The final refined model of AmbD contains two chains (A & B) in the asymmetric unit. With the exception of residues at the N- and C-terminus, chain A could be fully traced, whereas a gap of eight residues remained in chain B. Summary information about modeled residues in both chains are shown in table 4-10.

Table 4-10: Residues of AmbD that could be modeled at a resolution of 2.9 Å.

Chains	Start	End	Gaps
A	D19	H337	-
B	G18	H337	D92- A99

According to size-exclusion chromatography, AmbD was also expected to be a monomer in solution. The final structure of AmbD was submitted to PDBePISA (http://www.ebi.ac.uk/pdbe/pisa/), which did not suggest any specific interactions that could result in the formation of stable quaternary structures (Evgeny Krissinel, 2010; E. Krissinel & Henrick, 2007). Therefore, the presence of two monomers (Fig 4-18A) in the asymmetric unit was caused by crystal packing, and superposition of these two molecules reveals high structural identity (Fig 4-18B). The more complete chain A will be shown for the following analysis.

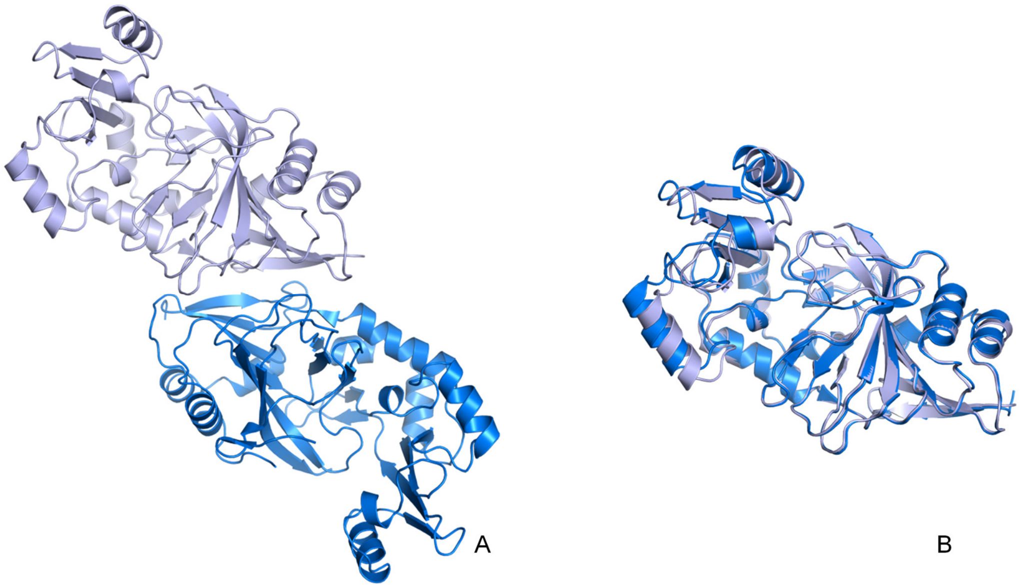


Fig 4-18: Overall structure of AmbD. A: two chains present in one asymmetric unit. Chain A is colored in lightblue and chain B in marine. B: superposition of the two chains.

4.1.12 Secondary and tertiary structure of AmbD

The structure of AmbD was submitted to PDBsum, which generated the information about secondary structure and topology shown in Fig 4-19 and Fig 4-20. The stereo view of overall structure was prepared with PyMOL (Fig 4-21). Each AmbD monomer contains 4 β-sheets (16 strands) and 14 α-helixes. Similar to AmbC 30-350, the core of AmbD consists of a jelly roll β-sandwich of anti-parallel strands. Near the core area, conserved residues His^{122} and His^{316} located on two anti-parallel β-strands respectively, and the third conserved residue Glu^{124} sit on one extended loop.

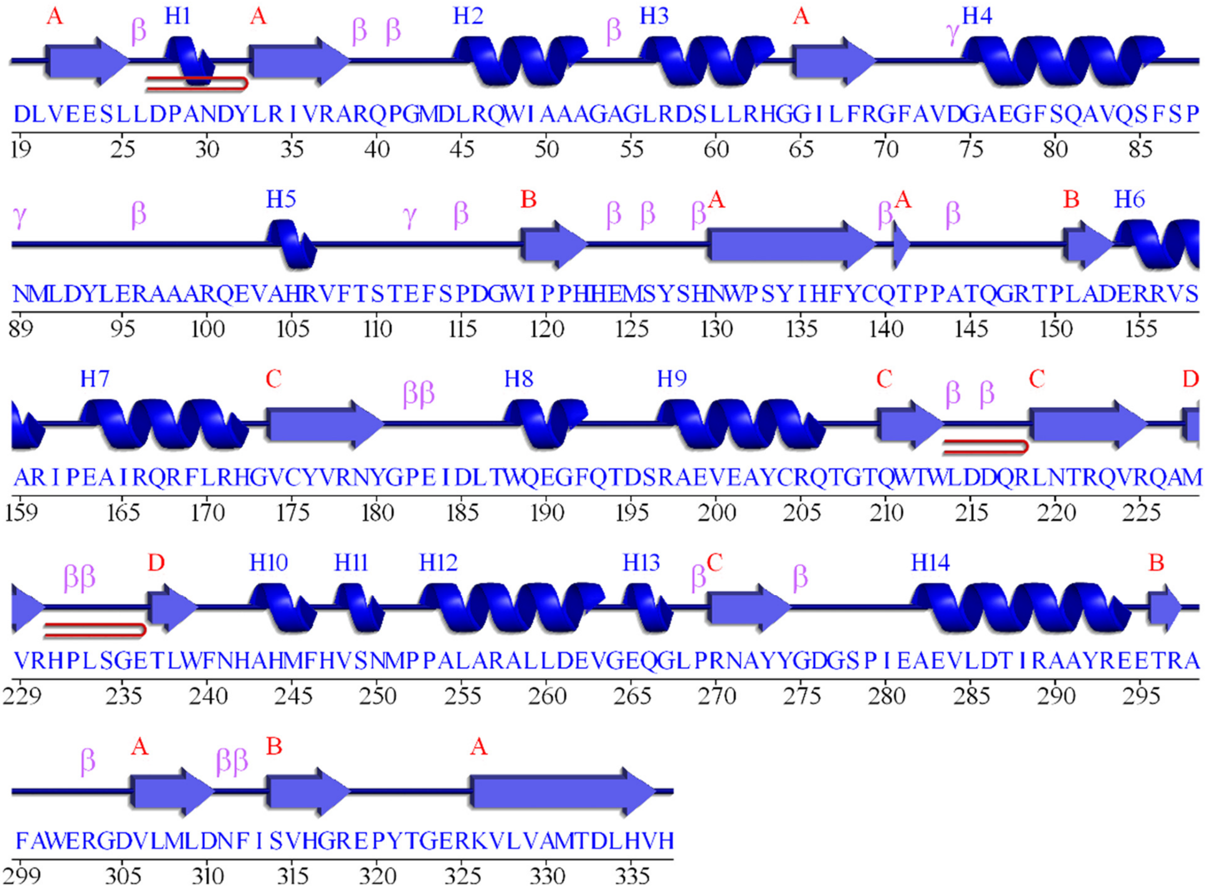


Fig 4-19: primary and corresponding secondary structure of AmbD. The 14 α-helixes are shown as blue letters H1-H14, and 4 β-sheets formed by continuous β-strands are shown as red letters A-D. β means beta turn, γ means gamma turn, red clip means beta hairpin.

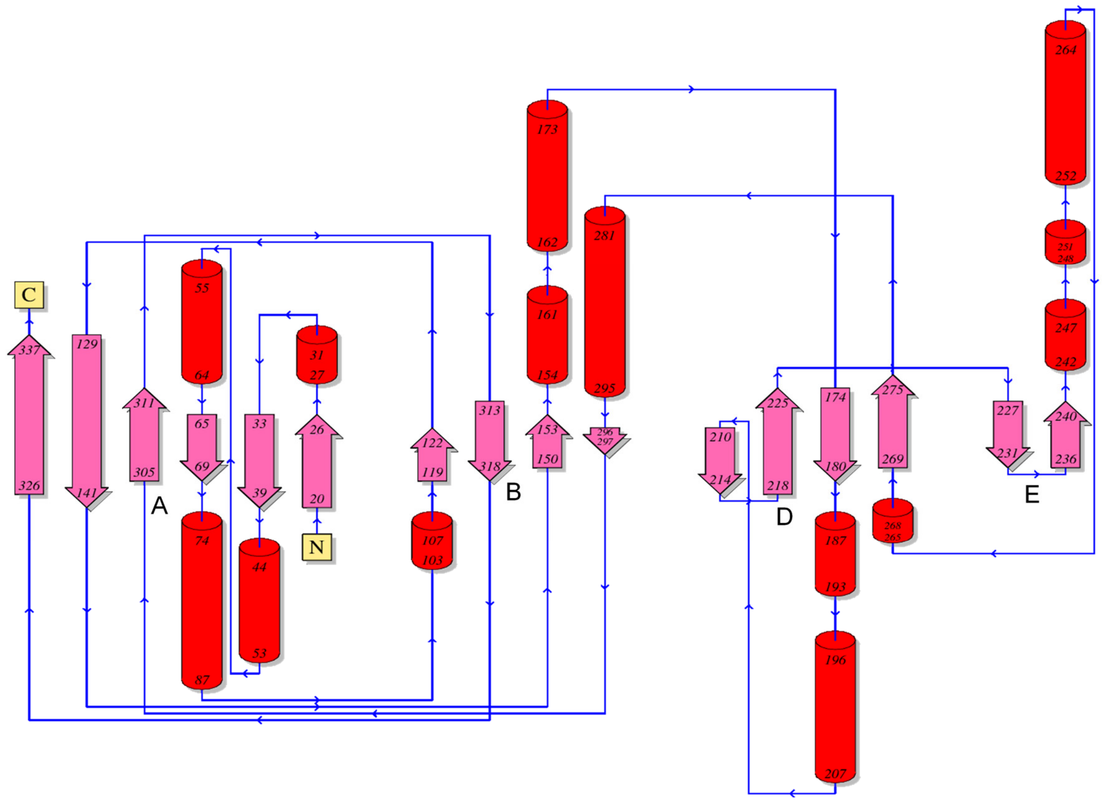


Fig 4-20: Topology of AmbD. Red columns represent 14 α-helixes, and pink arrows represent 16 β-strands classified to 4 β-sheets (A, B, D and E). The supposed conserved metal ion binding residues His^{122}, Glu^{124} and His^{316} are located near the B β-sheet.

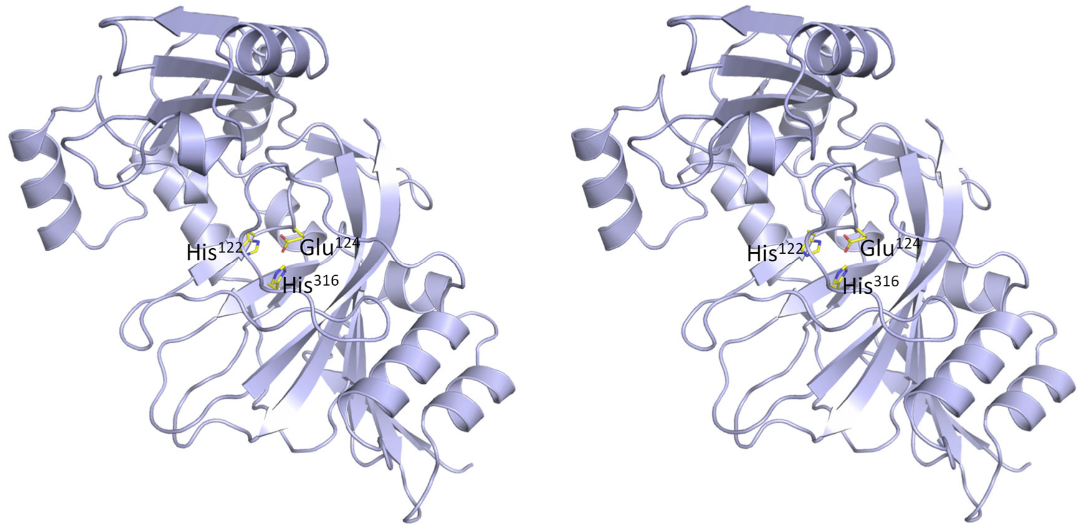


Fig 4-21: Stereo view of an AmbD monomer shown as cartoon colored in lightblue. Conserved residues His^{122}, Glu^{124} and His^{316} are shown as sticks colored by elements (carbon atoms are colored in yellow).

Furthermore, the active site metal was absent in the structure of apo AmbD (Fig 4-22), so the coordinated Glu^{124} was free to be contacted. Residue Arg^{100} located on the corresponding loop interacted with Glu^{124} to somehow pull the loop towards the active site (Fig 4-23).

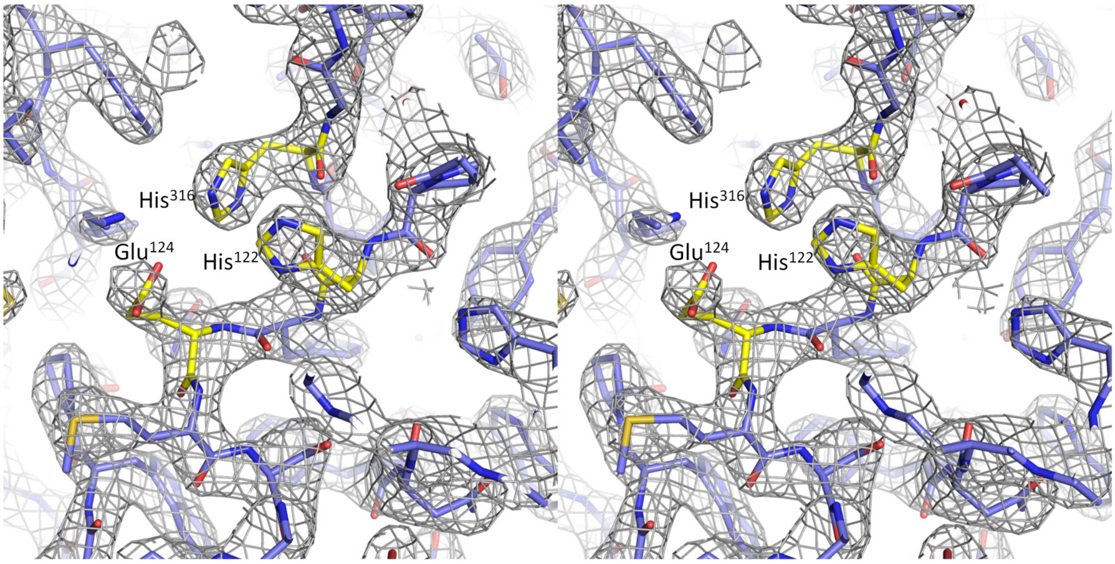


Fig 4-22: Stereo view of the active site of AmbD with electron density map. A 2F$_{obs}$ – F$_{calc}$ map (grey mesh) is calculated and contoured level at 1.5σ. The conserved residues are labeled and colored by elements (carbon atoms are colored in yellow).

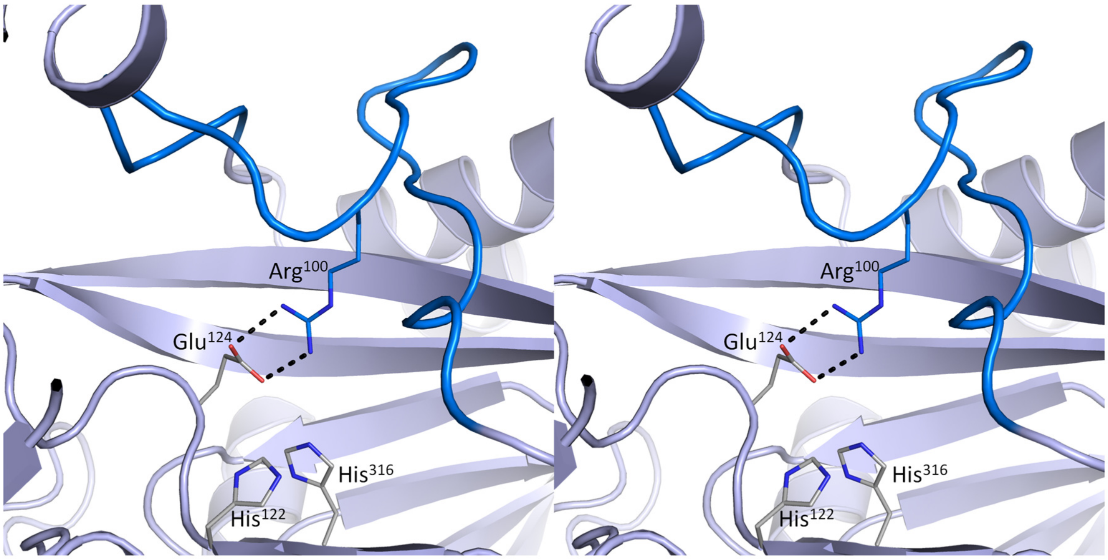


Fig 4-23: Stereo view of the active site of the structure of apo AmbD. The corresponding loop is colored in marine. Three conserved residues are shown as sticks colored by elements (carbon atoms are colored in grey). The residue Glu^{124} interacts with Arg^{100}

locating on the loop by hydrogen bonds. Overall structure of apo AmbD is shown in cartoon colored in lightblue.

4.1.13 Comparison between AmbC 30-350 and AmbD

The secondary structure and topology were a little bit different from that of AmbC 30-350 (Fig 4-10, 4-11, 4-19 and 4-20). At the N-terminus of AmbD, one extra α-helix and β-strand could be built, and the β-strand contributed to the first β-sheet A. The β-sheet C consists of two tiny strands located on the protein surface and was unique for AmbC 30-350, resulting 4 β-sheets in total for AmbD, one less than in AmbC 30-350. This β-sheet was far away from the centre, and the overall structures were very similar with an r.m.s.d. value of 1.943 Å, especially the core anti-parallel β-sheets which are responsible for the active site. For structure of AmbC 30-350, the core area was more closed than that of AmbD (Fig 4-24). This may be influenced by the binding of the metal.

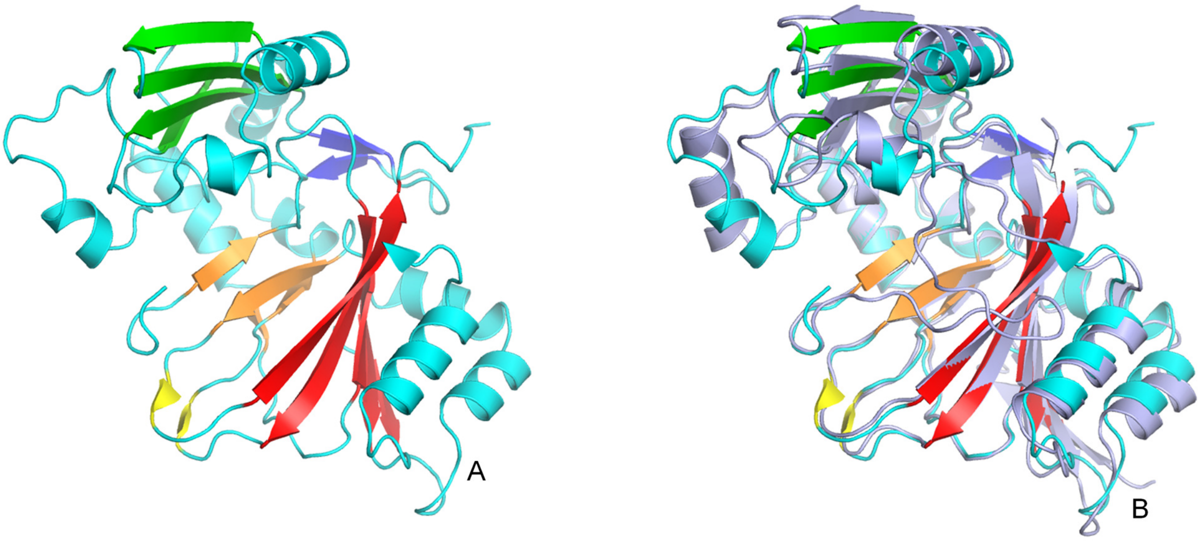


Fig 4-24: Structure of AmbC 30-350 and its superposition with AmbD. A: Structure of AmbC 30-350. Five β-sheets (Fig 4-11) are colored in red, orange, yellow, green and blue respectively. The α-helixes and loops are colored in cyan. B: Superposition of AmbC 30-350 and AmbD. The structure of AmbD is colored in lightblue. The conformations of the β-strands colored in red and orange mainly contributing to the active site are highly consistent.

The structure of AmbD was determined at lower resolution of 2.9 Å, and the metal near the active site was also missing. The loop above the metal binding site could nevertheless be built in chain A (Fig 4-25). As mentioned in the background part, flexible loops which are involved in substrate binding and closing the active site during reactions are a common feature of Fe^{2+}/α-ketoglutarate-dependent dioxygenases.

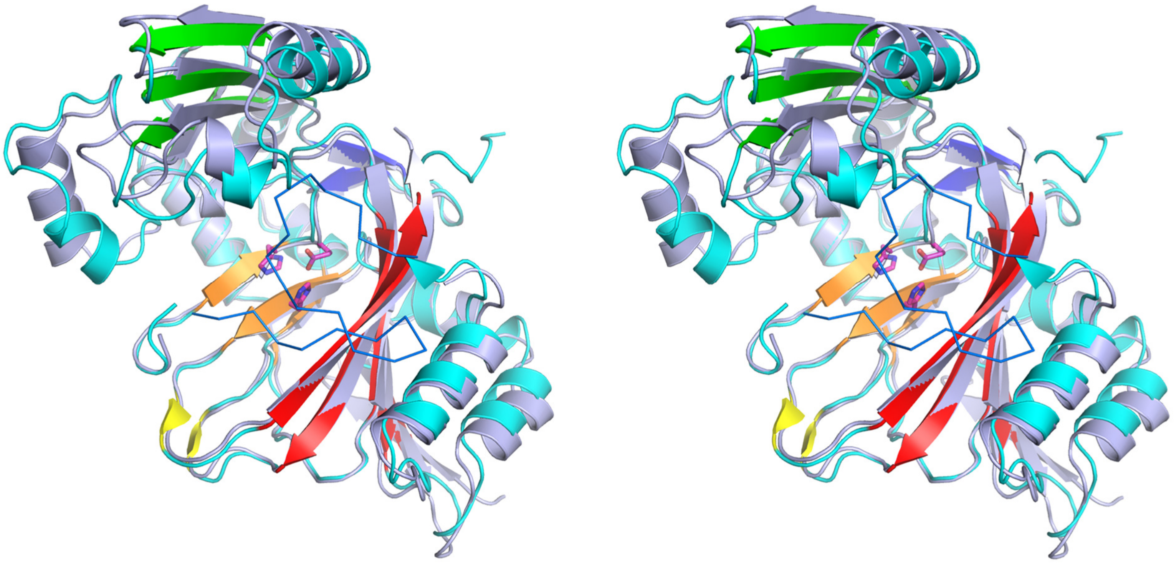


Fig 4-25: Stereo view of superposition of AmbC 30-350 and AmbD. The coloring is the same as in Fig 4-24. Conserved residues are shown as sticks and colored by elements (carbon atoms are colored in magenta). The flexible loop from the structure of AmbD is shown as ribbon and colored in marine.

4.2 Complex structure of AmbC with N-oxalylglycine (NOG)

4.2.1 The preparation of complex protein and crystallization

N-oxalylglycine (NOG, Fig 4-26 B), an inhibitor of α-ketoglutarate-dependent enzymes (Hausinger, 2004), was reported to bind Tet-like dioxygenase at the active site (Hashimoto *et al.*, 2014). A stock solution of NOG at 4 mM concentration was prepared, and mixed with 0.6 mM purified AmbC 30-350 at a volume ratio of 1:2, followed by incubation on ice for 3 hours for complex formation before setting up crystallization screens.

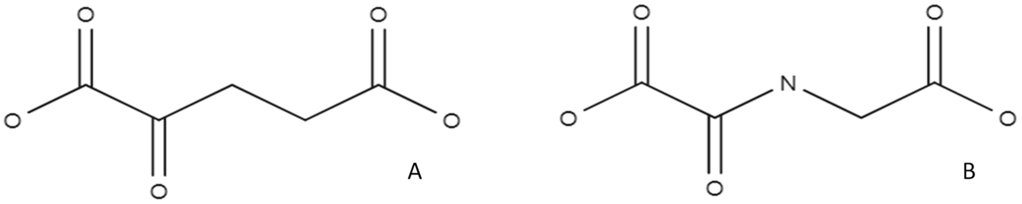


Fig 4-26: Structural formula of α-ketoglutarate (A) and N-oxalylglycine (B). The only difference is the β carbon atom in (A) that is replaced by nitrogen in (B).

The optimization kits prepared for apo AmbC 30-350 were tried first and well-shaped crystals emerged after seven days with a precipitant consisting of 0.0833 M HEPES pH 7.8, 0.172 M $MgCl_2$ and 31.1% PEG3350 (Fig 4-27).

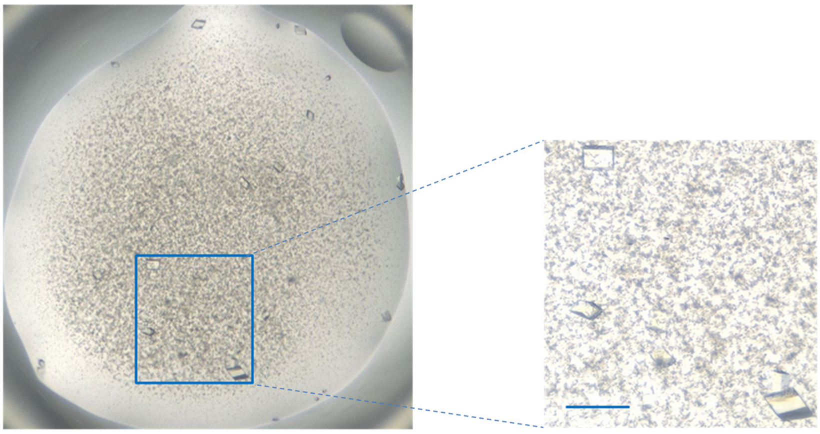


Fig. 4-27 Crystals of AmbC 30-350 in complexwith NOG obtained from optimized crystallization condition (0.0833 M HEPES pH 7.8, 0.172 M $MgCl_2$ and 31.1% PEG3350). Scale bar is 100 μm.

Given that the crystallization condition contains 31.1% PEG3350, there was no need for additional cryoprotection. Four crystals were flash frozen in liquid nitrogen and sent to beamline X06DA-PXIII in SLS (Waltersperger *et al.*, 2015) where four datasets were collected. 3600 images of 0.1° rotation were collected on a Pilatus 2M-F detector at a wavelength of λ = 1.00004 Å. Detailed statistics of the best dataset are shown in Table 4-11.

Table 4-11: Data collection statistics for AmbC 30-350 in complexwith NOG calculated by the aimless program (Evans, 2011; Evans PR, 2013). Values in parentheses are for the highest resolution shell.

Dataset	AmbC 30-350 in complexwith NOG
Wavelength (Å) / beamline	1.00004 / SLS, PXIII
Space group	I 222
Resolution range (Å)	38.44 (1.84-1.80)
Unit cell parameters (°)	90, 90, 90
(Å)	72.38, 76.89, 111.29
Mosaicity (°)*	0.353
Rpim	7.3 (39.7)
Total number of observations	351566 (17684)
Total number unique	29162 (1699)
Mean I / σ (I)	8.7 (2.1)
Mn (I) half-set correlation CC (1/2) (%)	99.1 (64.9)
Completeness (%)	100 (100)
Multiplicity	12.1 (10.4)

*Mosaicity value is from XDS (Kabsch, 2010)

The crystal of AmbC in complex with NOG diffracted to a resolution of 1.8 Å and reflections were indexed in space group I 222 with cell dimensions a = 72.38, b = 76.89, c = 112.29. There was one molecule in the asymmetric unit with 43.3% solvent according to Matthews_coef calculation in CCP4 (Kantardjieff & Rupp, 2003; Matthews, 1968).

4.2.2 Structure determination and refinement of AmbC 30-350 in complex with NOG

The structure of AmbC 30-350 in complexwith NOG was determined by molecular replacement using the structure of AmbC 30-350 as the search model. Refinement with phenix.refine (Headd *et al.*, 2014; Terwilliger, 2003) and manual adjustments in Coot (Emsley & Cowtan, 2004) followed to polish the final structure of the complex with R_{work} = 0.1957 and R_{free} = 0.2219 with 97.73% of the residues in Ramachandran-favored regions and a *MolProbity* (Chen *et al.*, 2010; Davis *et al.*, 2007) score of 1.20 (Table 4-12). The ligand NOG was added with Coot.

Table 4-12: Refinement statistics of AmbC 30-350 in complex with NOG generated with Phenix (Diederichs & Karplus, 2013). Values in parentheses are for the highest resolution shell.

Dataset	AmbC 30-350 in complexwith NOG
Resolution range (Å)	38.44 - 1.8 (1.864-1.8)
Total reflections	58319 (5744)
Unique reflections	29161 (2873)
Reflections used in refinement	29156 (2873)
Reflections used for R-free	1377 (143)
R_{work}	0.1957 (0.2474)
R_{free}	0.2219 (0.2790)
Number of non-hydrogen atoms	2709
macromolecules	2499
Ligands	14
Solvent	196
Protein residues	310
RMS(bonds)	0.004
RMS(angles)	1
Ramachandran favored (%)	97.73
Ramachandran outliers (%)	0

Rotamer outliers (%)	0.39
Clashscore	3.45
Average B-factor	19.12
macromolecules	19.12
Ligands	15.74
Solvent	19.37
MolProbity score*	1.20

*MolProbity value is from MolProbity (http://molprobity.biochem.duke.edu/) (Chen *et al.*, 2010; Davis *et al.*, 2007)

4.2.3 Overall structure of AmbC 30-350 in complex with NOG

The final refined complex structure contains one molecule in the asymmetric unit. Ten amino acids located in N- and C-terminus are missing. All the residues that could be modeled are shown in table 4-13, and the overall structure of AmbC 30-350 in complex with NOG was shown in Fig 4-28. The active-site nickel (see analysis from CheckMyMetal server in appendix Fig 6-1) and the co-substrate analog NOG could also be built (Fig 4-29).

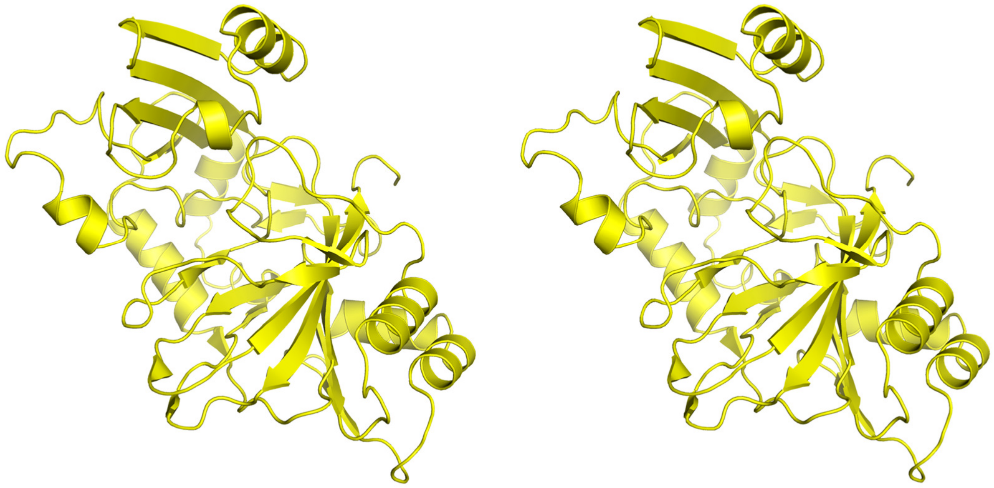


Fig 4-28: Stereo view of the complex structure of AmbC 30-350 with NOG shown as cartoon colored in yellow.

Table 4-13: Residues of AmbC 30-350 in complexwith NOG that can be modeled at the resolution of 1.8 Å. For clarity, the first residue of the truncated AmbC 30-350 is numbered 1.

Chain	Start	End	Gap
A	R6	L315	-

4.2.4 The active site of AmbC 30-350 in complex with NOG

The NOG ligand was well defined in the electron density maps (Fig 4-29). It was added with coot and refinement was followed with Phenix. The central nickel was coordinated by the same three residues, His^{99}, Glu^{101} and His^{293}, as in the structure of apo AmbC 30-350.

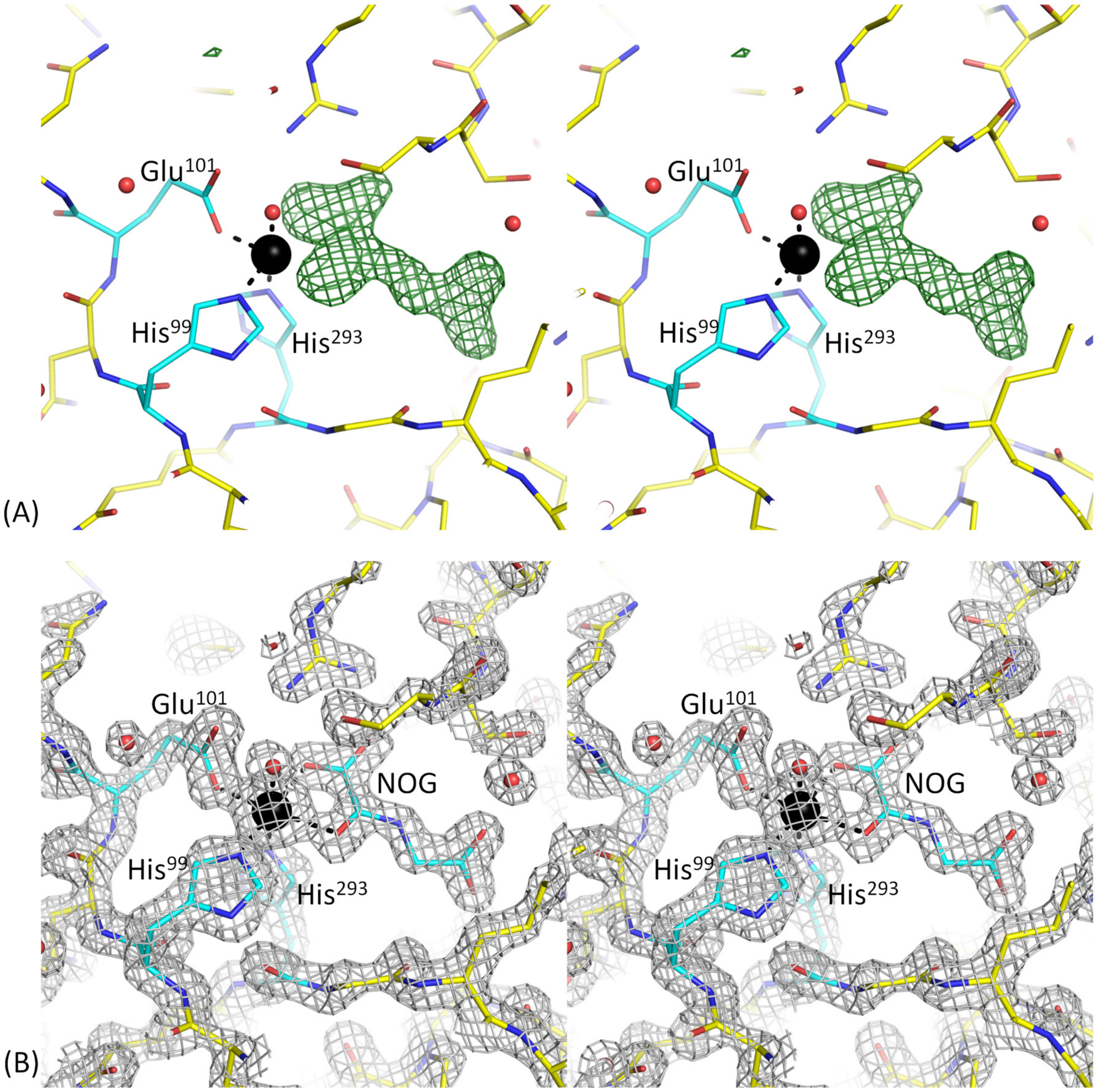


Fig 4-29: Stereo view of the active site of AmbC 30-350 in complexwith NOG and electron density maps. (A): $F_{obs} - F_{calc}$ difference electron density map (green mesh) contoured level at 4.0 σ, with the metal nickel (black sphere) and one coordinated water molecule. (B): The $2F_{obs} - F_{calc}$ map (grey mesh) calculated with NOG contoured level at 2.0 σ. The water molecules are shown as spheres colored in red. Three coordinated residues His^{99}, Glu^{101} and His^{293} and the ligand NOG are labeled and colored by elements (carbon atoms are colored in cyan).

Two coordinated water molecules in the apo structure were replaced by the keto-group and the carboxyl oxygen atoms of NOG. The last coordination site occupied by a water molecule represents a putative oxygen binding site, which will recruit the oxygen after the specific substrate is introduced (Fig 4-30). The ligand NOG has binding interactions to the residues Arg^{77}, Thr^{88}, Thr^{126} and Arg^{302} which are about 2.4-3.0 Å apart from the NOG molecule (Fig 4-30 and Fig 4-31).

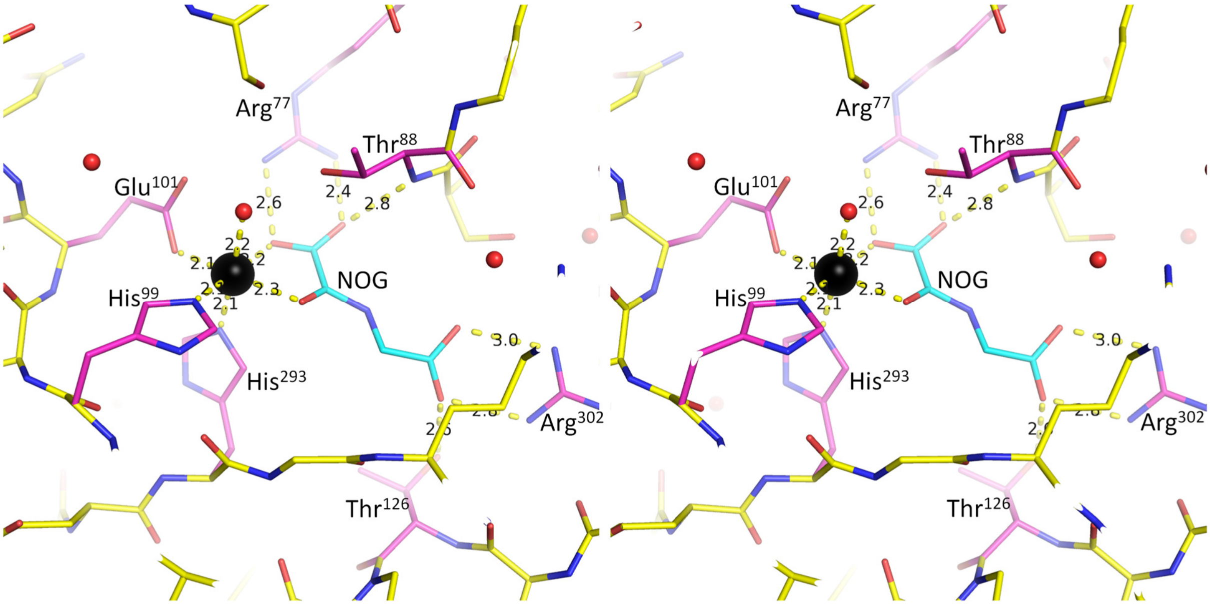


Fig 4-30: Stereo view of the active site of AmbC 30-350 in complexwith NOG. The distances between nickel atom and all coordinated ligand atoms and that of NOG are measured with PyMOL. Besides the three conserved residues (Fig 4-29), the NOG coordinating residues Arg^{77}, Thr^{88}, Thr^{126} and Arg^{302} are also colored by elements (carbon atoms are colored in magenta).

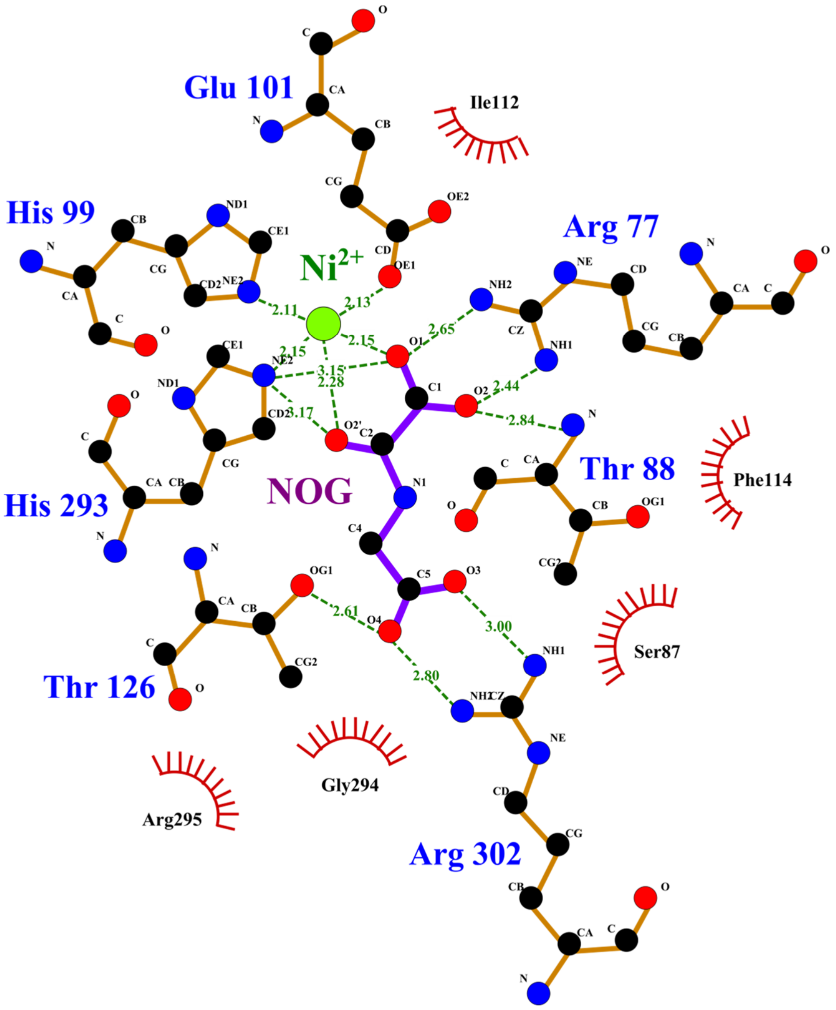


Fig 4-31: Arrangement of the active site of AmbC 30-350 in complexwith NOG. The central nickel atom is coordinated by three conserved residues and the keto-group and an carboxyl oxygen of NOG. A water molecule occupied the last coordination site which represents a putative oxygen binding site. Except binding with nickel, the ligand NOG interacts with other four residues Arg^{77}, Thr^{88}, Thr^{126} and Arg^{302} by hydrogen bonds. Residues involved in non-ligand hydrophobic interactions are also shown.

4.2.5 Structural comparisons of NOG-bound AmbC 30-350 with apo one

4.2.5.1 Superposition with the structure of apo AmbC 30-350

For the structure of apo AmbC 30-350, two molecules were defined in the asymmetric unit. Both of them were not intact, resulting in a gap (around 20 amino acids) in between (Table 4-6). Because of high similarity (Fig 4-9), chain B had been chosen to perform superposition with the complex structure of AmbC 30-350 with NOG. Overall the architectures are almost the same, and the most significant difference is that the missing loop in the structure of apo AmbC 30-350 could be built in the complex structure (Fig 4-32A). This loop is located exactly above the active site, suggesting that it plays the role of a switch to regulate substrate binding (Fig 4-32B).

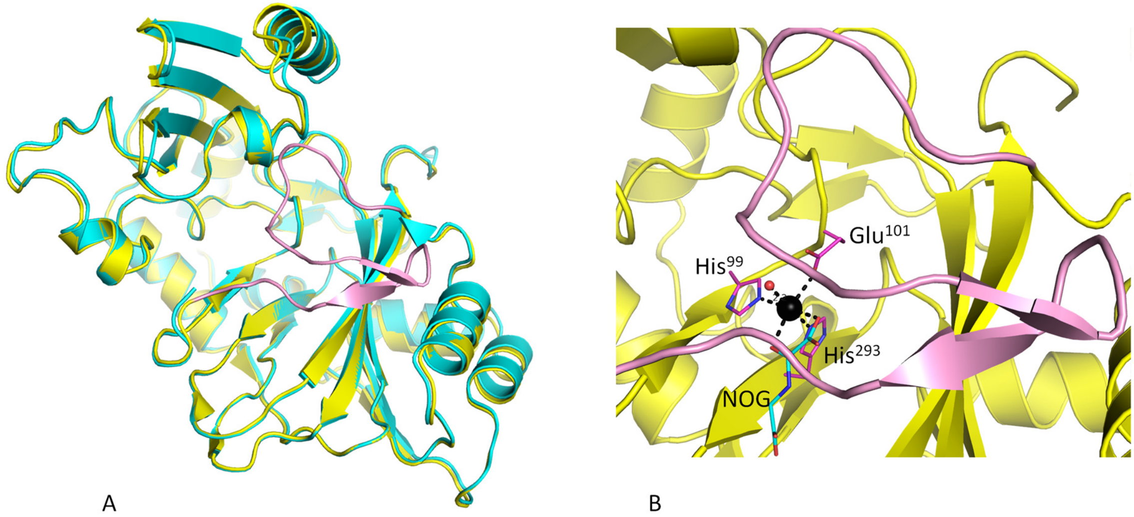


Fig 4-32: (A): Superposition between structure of apo AmbC 30-350 and the complex one bound NOG. Apo structure is colored in cyan. The complex one is colored in yellow, of which the flexible loop that was absent in the structure of apo AmbC 30-350 is colored in pink. (B): The active site of complex structure AmbC 30-350 with NOG. Nickel is shown as sphere colored in black. The conserved residues His^{99}, Glu^{101} and His^{293} are shown as sticks colored by elements (carbon atoms are colored in magenta). NOG is shown as sticks colored by elements (carbon atoms are colored in cyan). One water molecule coordinated with nickel is shown as sphere colored in red.

4.2.5.3 B-factor analysis

The B-factor can indicate relative vibrational motion of different parts of structures. It is measured in units of Å^2. Residues with low B-factor belong to a well organized part of one structure, while those with high B-factor generally belong to more flexible regions.

The better built chain B of the structure of apo AmbC 30-350 showed three distinct α-helixes with relatively high B-factor up to a maximum of 71 Å^2 (Fig 4-33A). Among them, helix 2 had the highest B-factor, and together with helix 3, it establishes the largest flexible part of the structure of apo AmbC 30-350. Helix 1 has lower B-factors and is located further away from the active site. The B-factors of AmbC 30-350 in complexstructure with NOG were then analyzed (Fig 4-33B). The corresponding locations of helix 1 and helix 2 showed well organized trends, which suggested that both helixes may play a role of regulating substrate binding. The missing loop could be built in the complex structure, and it showed the largest B-factor of 56 Å^2 in this structure. Because of its special location and flexibility, it seems likely that the loop may involve in substrate binding and closing off the active site to prevent the release of reactive oxygen species (ROS).

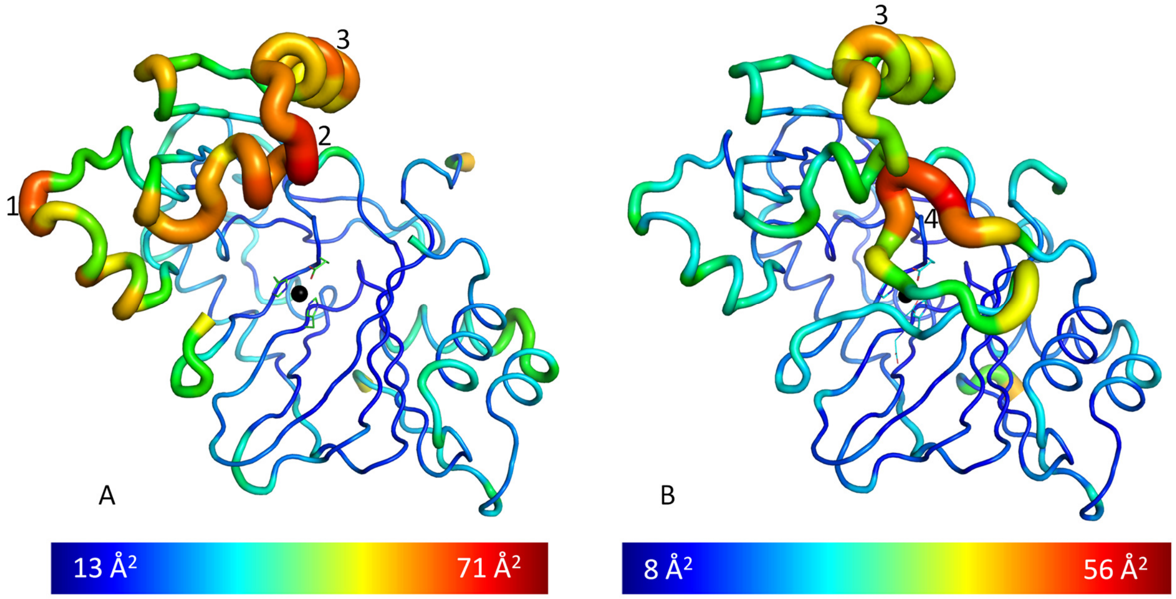


Fig 4-33: B-factor distribution in the structure of apo AmbC 30-350 and the complex one. Wider ribbon and warmer colors represent higher B-factors. (A): For the apo AmbC 30-350, three flexible α-helixes near the active site could be identified and labeled 1, 2 and 3. (B): For the complex one, the flexible α-helix 1 and 2 belong to the structure of apo AmbC 30-350 return to organized state. Newly built loop 4 shows the highest B-factor in this complex structure. The active site metal is shown as black sphere and conserved residues are shown as sticks colored by elements (carbon atoms are colored in cyan).

4.3 Complex structure of AmbD with α-ketoglutarate (α-KG)

4.3.1 Purification and crystallization of AmbD in complex with α-KG

The metal ion near the active site could not be defined in the structure of apo AmbD, such that it was necessary to incubate the protein with metal before introducing co-substrate α-ketoglutarate (α-KG) inside. Cobalt and manganese share similar properties with the iron, and ion states of them (Co^{2+} and Mn^{2+}) are more stable than ferrous. Cobalt was chosen to replace the iron coordinating in the structure of AmbD in this study.

Protein AmbD was purified as usual. After standard Ni^{2+}-IMAC and SEC purifications, around 5 ml AmbD sample at 0.25 mM concentration was obtained, and incubations with 120 µl 100 mM Co^{2+} and 120 µl 100 mM α-KG on ice sequentially followed. Then, a second SEC step was used to remove extra metals and α-KG. The result of SDS-PAGE for SEC was shown in Fig 4-34. The purest fractions were collected and concentrated to 17 mg/ml for further crystallization screens.

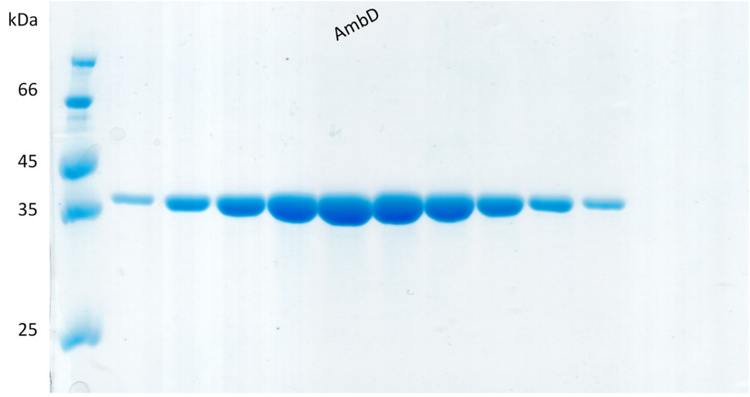


Fig 4-34: SDS-PAGE followed by InstantBlue staining for protein complex AmbD with Co^{2+} and α-KG.

Initial crystallization condition screens were set up for this complex. Precipitation was observed in most of the drops at first, but after 87 days perfect looking crystals were found with a precipitant consisting of 0.123 M HEPES pH 7.8, 0.15 M $MgCl_2$ and 28.3%

PEG4000. These crystals emerged from a drop that contained precipitate earlier (Fig 4-35).

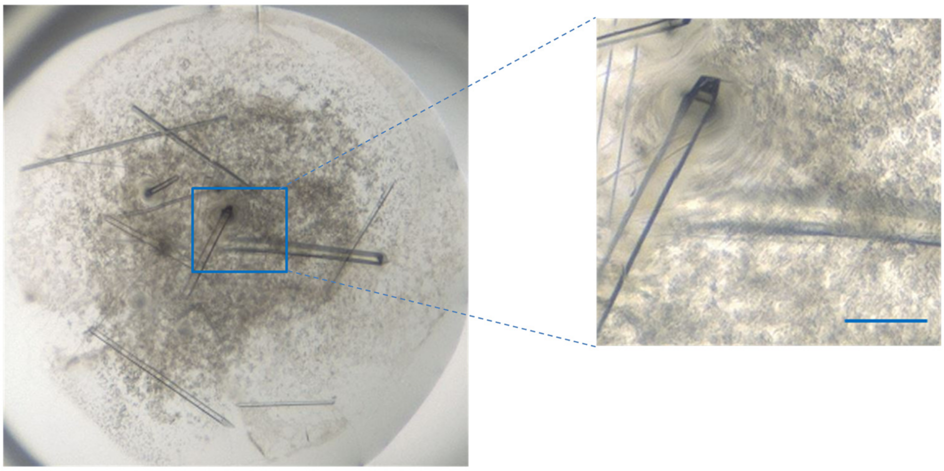


Fig 4-35: Crystals of AmbD incubated with Co^{2+}/α-KG obtained from the optimization condition for AmbC 30-350 (0.123 M HEPES pH 7.8, 0.15 M $MgCl_2$ and 28.3% PEG4000). Scale bar is 100 µm.

Three crystals were sent to beamline P11 in DESY (Burkhardt, 2016), and 3600 images of 0.1° rotation were collected on a Pilatus 6M detector at a wavelength of λ = 1.0332 Å. The best dataset had been analyzed and detailed statistics are shown in Table 4-14.

Table 4-14: Data collection statistics for AmbD in complex with α-KG generated with aimless program (Evans, 2011; Evans PR, 2013). Values in parentheses are for the highest resolution shell.

Dataset	AmbD in complex with α-KG
Wavelength (Å) / beamline	1.0332 / DESY, P11
Space group	$P2_12_12_1$
Resolution range (Å)	41.82-1.6 (1.63-1.60)
Unit cell parameters (°)	90, 90, 90
(Å)	49.53, 71.49, 78.03
Mosaicity (°)*	0.17
Rpim	2.8 (27.8)
Total number of observations	480885 (23002)
Total number unique	36769 (1767)
Mean I / σ (I)	20.5 (2.8)

Mn (I) half-set correlation CC (1/2) (%)	99.9 (71.9)
Completeness (%)	99.1 (97.0)
Multiplicity	13.1 (13.0)

*Mosaicity value is from XDS (Kabsch, 2010)

The crystal of AmbD in complex with α-KG diffracted to a resolution of 1.6 Å and reflections were indexed in space group $P2_12_12_1$ with cell dimensions a = 49.53, b = 71.49, c = 78.03. There was one molecule in the asymmetric unit according to Matthews_coef calculation in CCP4 (Kantardjieff & Rupp, 2003; Matthews, 1968).

4.3.2 Structure determination and refinement of AmbD in complex with α-KG

The structure of AmbD in complex with α-KG was determined by molecular replacement using the chain A from the structure of apo AmbD as the search model. The final structure was polished as usual with R_{work} = 0.1654 and R_{free} = 0.1896 and 95.92% of the residues in Ramachandran-favored regions and a *MolProbity* (Chen *et al.*, 2010; Davis *et al.*, 2007) score of 1.63 (Table 4-15). The ligand α-KG was added with Coot.

Table 4-15: Refinement statistics of AmbD in complex with α-KG generated with Phenix (Diederichs & Karplus, 2013). Values in parentheses are for the highest resolution shell.

Dataset	AmbD in complex with α-KG
Resolution range (Å)	41.82-1.602 (1.66-1.602)
Total reflections	482476 (47204)
Unique reflections	36723 (3552)
Reflections used in refinement	36718 (3552)
Reflections used for R-free	1702 (151)
R_{work}	0.1654 (0.2458)
R_{free}	0.1896 (0.2707)
Number of non-hydrogen atoms	2793
macromolecules	2594
ligands	11

solvent	188
Protein residues	321
RMS(bonds)	0.008
RMS(angles)	1.28
Ramachandran favored (%)	95.92
Ramachandran outliers (%)	0
Rotamer outliers (%)	1.49
Clashscore	4.33
Average B-factor	22.37
macromolecules	22.12
ligands	21.46
solvent	25.88
MolProbity score*	1.63

*MolProbity value is from MolProbity (http://molprobity.biochem.duke.edu/) (Chen *et al.*, 2010; Davis *et al.*, 2007)

4.3.3 Overall structure of AmbD in complex with α-KG

The final refined complex structure contains one molecule in the asymmetric unit. 24 amino acids located in N- and C-terminus are missing. All the residues that could be modeled are shown in table 4-16. The active-site metal cobalt (see analysis from CheckMyMetal server in appendix Fig 6-2) and co-substrate α-KG were built in chain B and chain C, respectively (Fig 4-36).

Table 4-16: Residues of AmbD in complex with α-KG that can be modeled at the resolution of 1.6 Å.

Chain	Start	End	Gap
A	D 19	P339	-

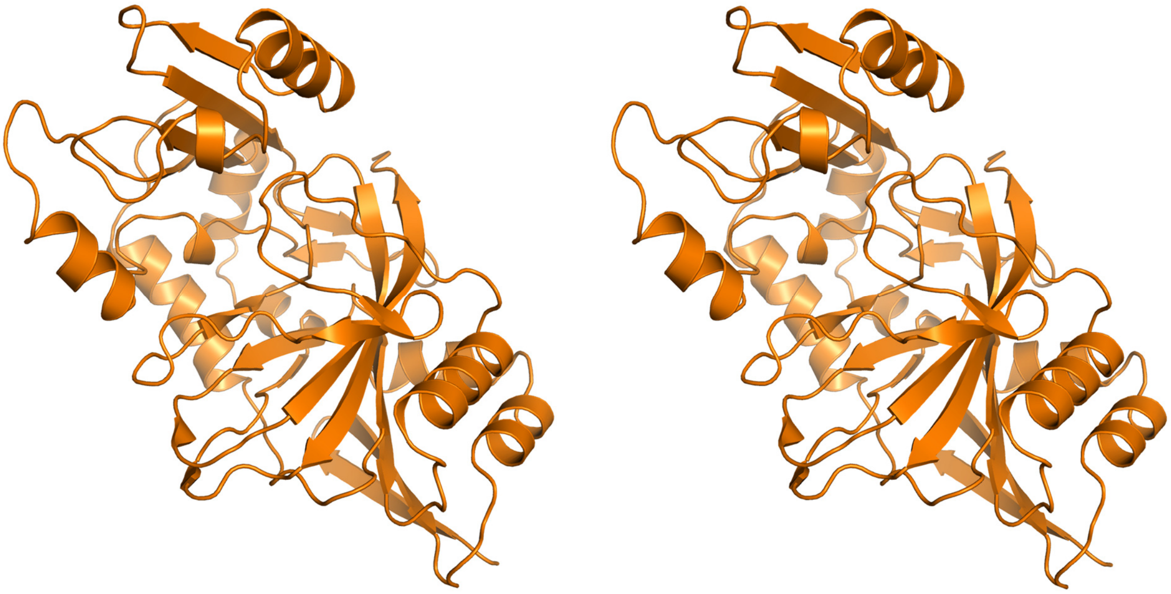


Fig 4-36: Stereo view of the complex structure of AmbD with α-KG shown as cartoon colored in orange.

4.3.4 The active site of AmbD bound with α-KG

The co-substrate α-KG was well defined in the electron density maps (Fig 4-37). It was added with coot and refinement followed with Phenix. The central cobalt was coordinated by the same three corresponding residues, His^{122}, Glu^{124} and His^{316}, as in the structure of AmbC 30-350. Two coordinated water molecules were replaced by the keto-group and the carboxyl oxygen atoms of α-KG. The last coordination site occupied by a water molecule represents a putative oxygen binding site, which will recruit the oxygen after the specific substrate is introduced (Fig 4-38). According to the α-KG, other binding interactions were conducted by residues Arg^{100}, Thr^{111}, Thr^{149} and Arg^{325} which were about 2.6-3.3 Å apart from the α-KG molecule, and H_2O^{178} here forms hydrogen bonds with α-KG and Glu^{112} respectively, together with an extra Ser^{110}-formed hydrogen bond with α-KG, which may strengthen this binding (Fig 4-38 and Fig 4-39).

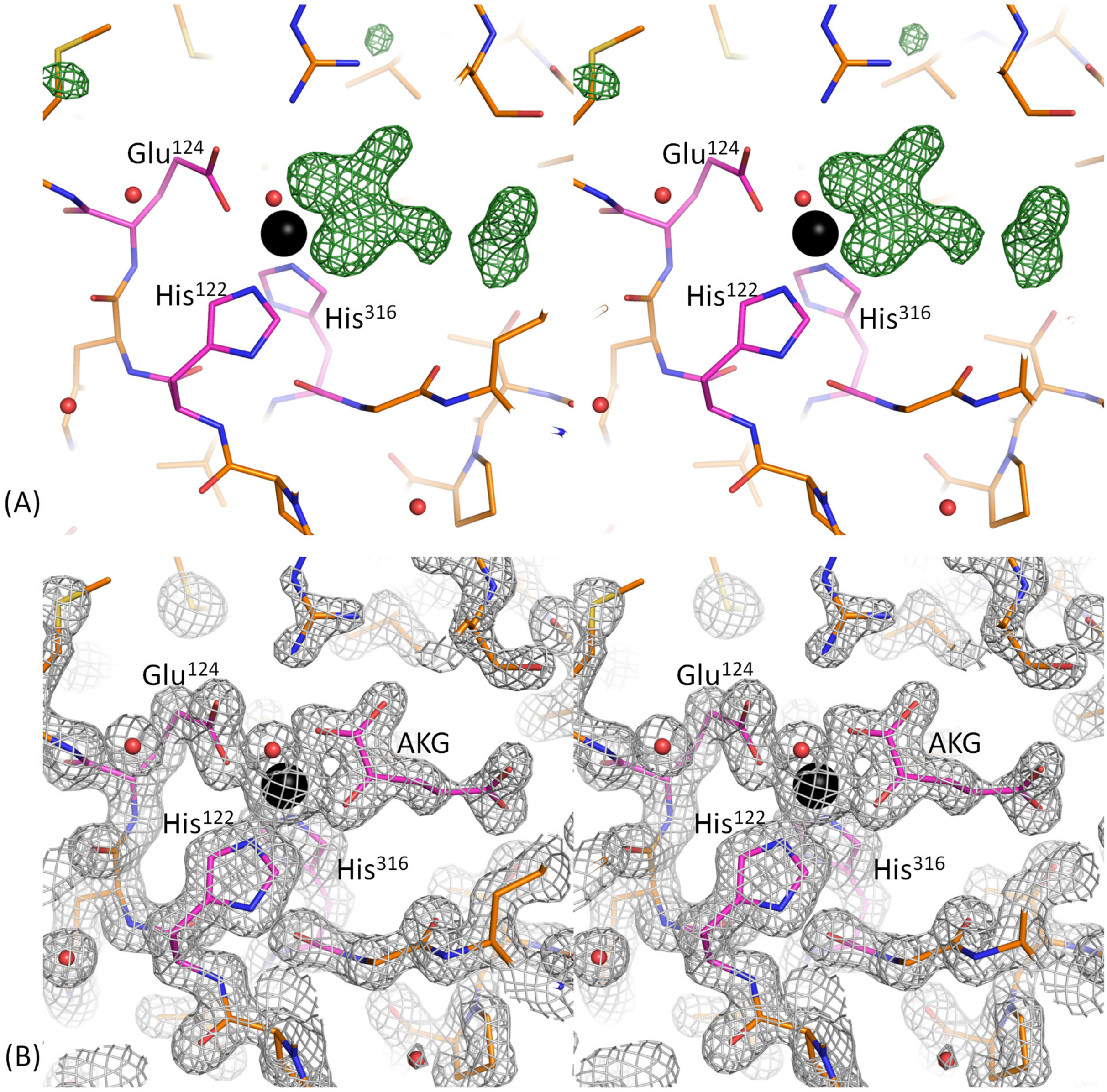


Fig 4-37: Stereo view of the active site of AmbD in complex with α-KG and electron density maps. (A): $F_{obs} - F_{calc}$ difference electron density map (green mesh) contoured level at 4.0 σ, with the metal cobalt (black sphere) and one coordinated water molecule. (B): The $2F_{obs} - F_{calc}$ map (grey mesh) calculated with α-KG contoured level at 1.5 σ. The water molecules are shown as spheres colored in red. Three coordinated residues His^{122}, Glu^{124} and His^{316} and the ligand α-KG are labeled and colored by elements (the carbon atoms belong to residues are colored in magenta, and α-KG in cyan).

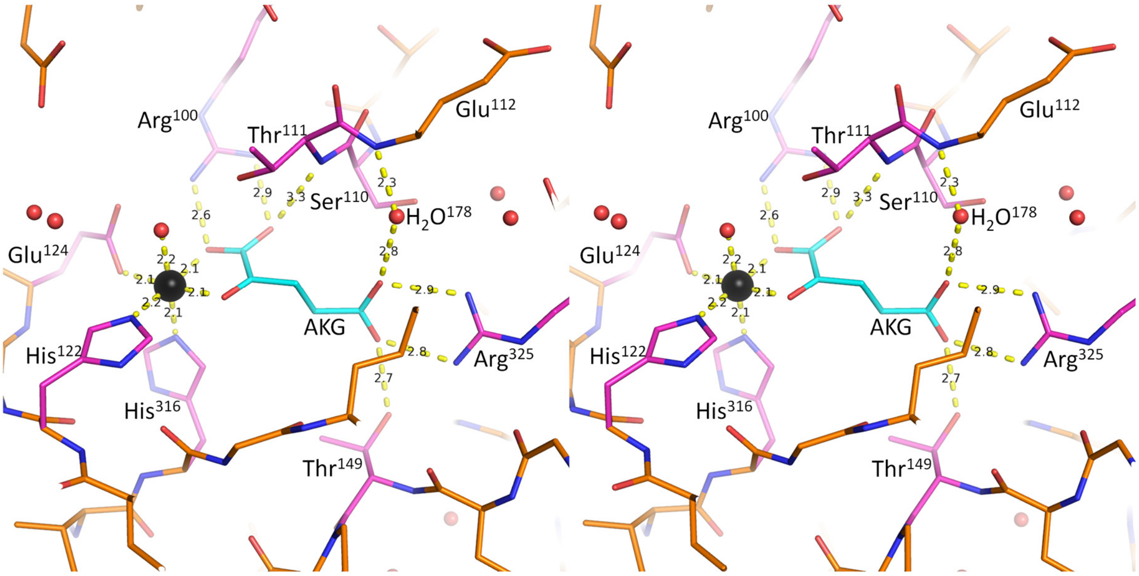


Fig 4-38: Stereo view of the active site of AmbD in complex with α-KG. The distances between cobalt atom and all coordinated ligand atoms and that of α-KG were measured with PyMOL. Besides the three conserved residues, the α-KG coordinating residues Arg^{100}, Thr^{111}, Thr^{149} and Arg^{325} were colored by elements (carbon atoms were colored in magenta). H_2O^{178} can form hydrogen bonds with α-KG (carbon atoms were colored in cyan) and Glu^{112} (carbon atoms were colored in orange) respectively.

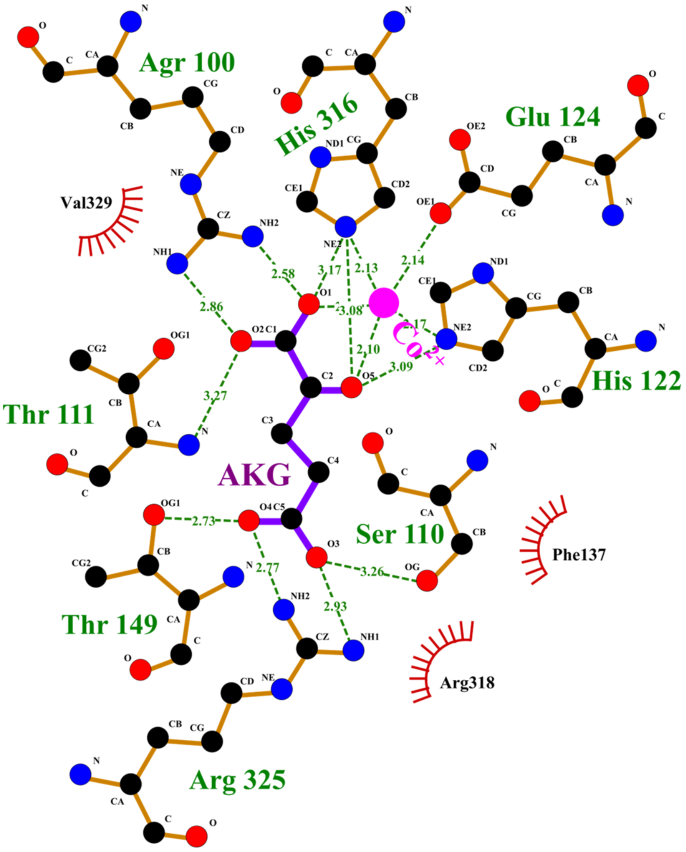


Fig 4-39: Arrangement of the active site of AmbD in complex with α-KG. The central cobalt atom was coordinated by three reserved residues and the keto-group and carboxyl oxygen of α-KG. Except binding with cobalt, ligand α-KG interacted with other five residues Arg^{100}, Thr^{111}, Thr^{149}, Arg^{325} and Ser^{100} by hydrogen bonds. Residues involved in non-ligand hydrophobic interactions are also shown.

4.3.5 Structural comparisons of α-KG-bound AmbD with apo one

4.3.5.2 Superposition with the structure of apo AmbD

For the structure of apo AmbD, the asymmetric unit contained two molecules of which the better defined chain A (Table 4-10) was chosen to perform the following superposition.

Structural superposition showed high similarity between apo AmbD and the complex with α-KG, except for one α-helix (Fig 4-40). Helix α10 locating in the structure of AmbD in complex with α-KG has moved towards the active site obviously in contrast with the one in the structure of apo AmbD (Fig 4-41A). This is a strong indication that the active site closed upon binding metal, co-substrate or/and specific substrate.

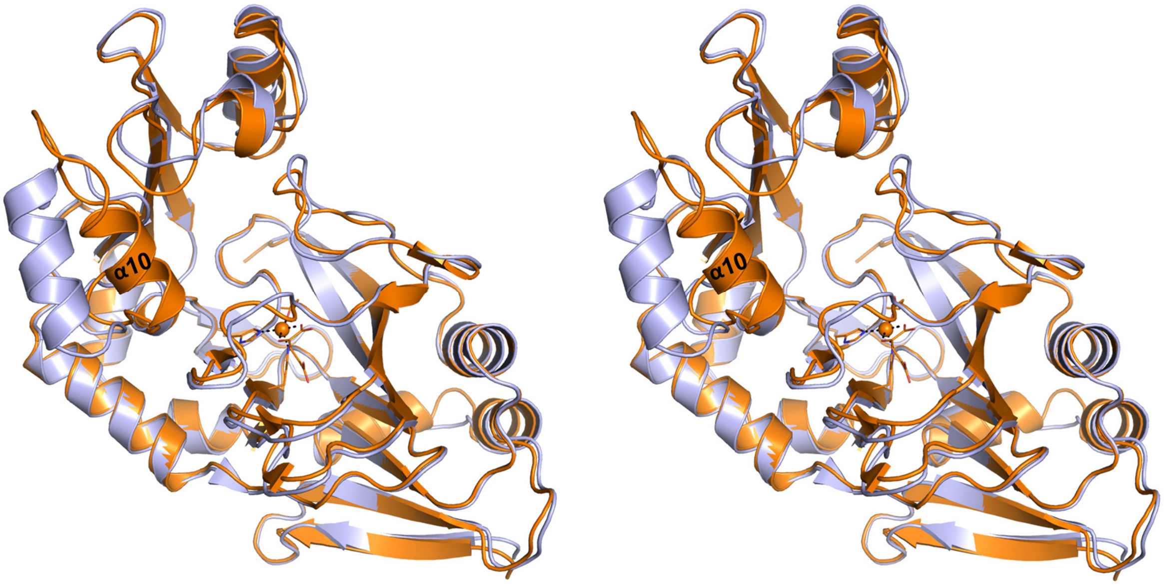


Fig 4-40: Stereo view of superposition between apo AmbD and complex one with α-KG. apo AmbD is colored in lightblue and complex one in orange. Helix α10 belongs to the complex structure is labeled. Cobalt is shown as sphere colored in orange, and the coordinated conserved residues and α-KG are shown as sticks colored by elements (carbon atoms are colored in orange).

Four residues were observed to conduct the binding between helix α10 and the flexible loop which was located above the active site. Ala^{254} and Arg^{257} located on α10 helix form hydrogen bonds with Ser^{114} and Asp^{116} located on a nearby loop, respectively (Fig 4-41B), which stabilizes the closed state.

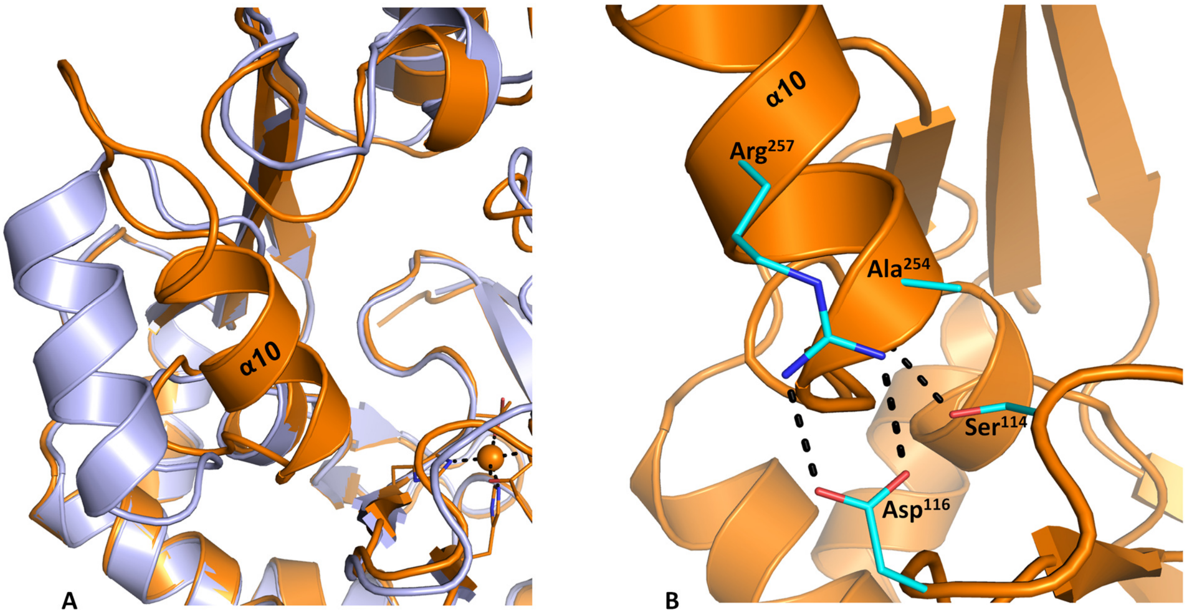


Fig 4-41: (A): Zoomed view of the helix α10. This helix from the structure of AmbD in complex with α-KG significantly moves towards the active site. Coloring information are the same as Fig 4-40. (B): Residues interaction between α10 helix and the flexible loop which is located above the active site. Ala^{254} and Arg^{257} formed hydrogen bonds with Ser^{114} and Asp^{116}, respectively. Involved residues are shown as sticks colored by elements (carbon atoms are colored in cyan).

4.3.5.3 B-factor analysis

The better built chain A of the structure of apo AmbD revealed two α-helices (labeled 1 and 2) and one loop (labeled 3) with relatively high B-factor up to a maximum of 95 Å^2 (Fig 4-42A). B-factor of AmbD complex structure with α-KG was then analyzed (Fig 4-42B). Accordingly, helix 1 and 2 are much less flexible in the complex, which is very similar to the structure of AmbC 30-350 in complex with NOG (Fig 4-33). The only one flexible part is the loop which is located above the active site, and it shows the highest B-factor of 102 Å^2 in this complex structure.

Together, these analyses show that binding of the metal ion and of α-KG or the co-substrate analog NOG lead to a closing of the active site, which is accompanied by ordering of helices 1 and 2 in AmbC and AmbD.

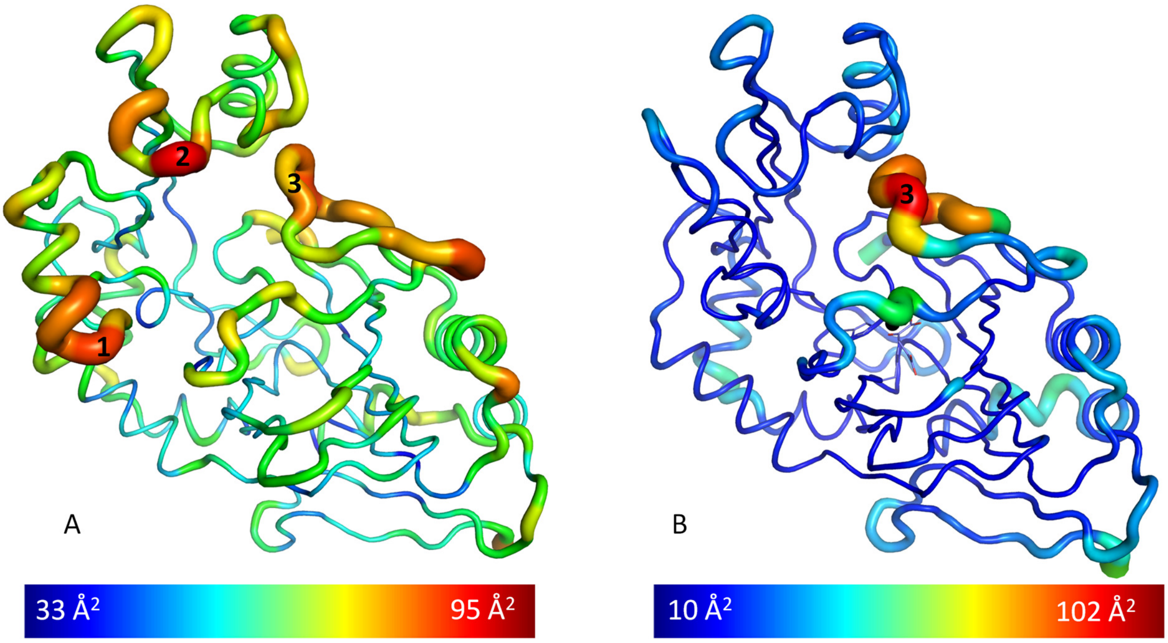


Fig 4-42: B-factor distribution in the structure of apo AmbD and the complex one with α-KG. Wider ribbon and warmer colors represent higher B-factors. (A): Flexible helices 1 and 2 and loop 3 are identified in the structure of apo AmbD. (B): In comparison to apo-AmbD, the flexible helix 1 and helix 2 assume a more ordered structure in the complex. The loop 3 shows the highest B-factor in this structure. The active site metal is shown as sphere colored in black. Conserved residues and co-substrate α-KG are shown as sticks (carbon atoms are colored in lightblue).

4.3.5.4 Surface analysis

The characteristics of residues near the active site can provide more hints to analyze how and why a substrate binds there. The proposed substrate for the two Fe/α-KG-dependent enzymes AmbC and AmbD is *L*-Glu, *L*-Glu-*L*-Ala or tripeptide *L*-Ala-*L*-Glu-*L*-Ala, and all modifications should happen on the core *L*-Glu moiety, which is a hydrophilic and negatively charged amino acid.

The electrostatic potential calculated by APBS plugin in PyMOL (Baker *et al.*, 2001) showed several positive and negative patches on the surface of AmbD (Fig 4-43). Near the active cavity, two residues Arg^{178} and Arg^{270} contributed a highly positively charged area, which may play a key role of recruiting the negatively charged *L*-Glu moiety.

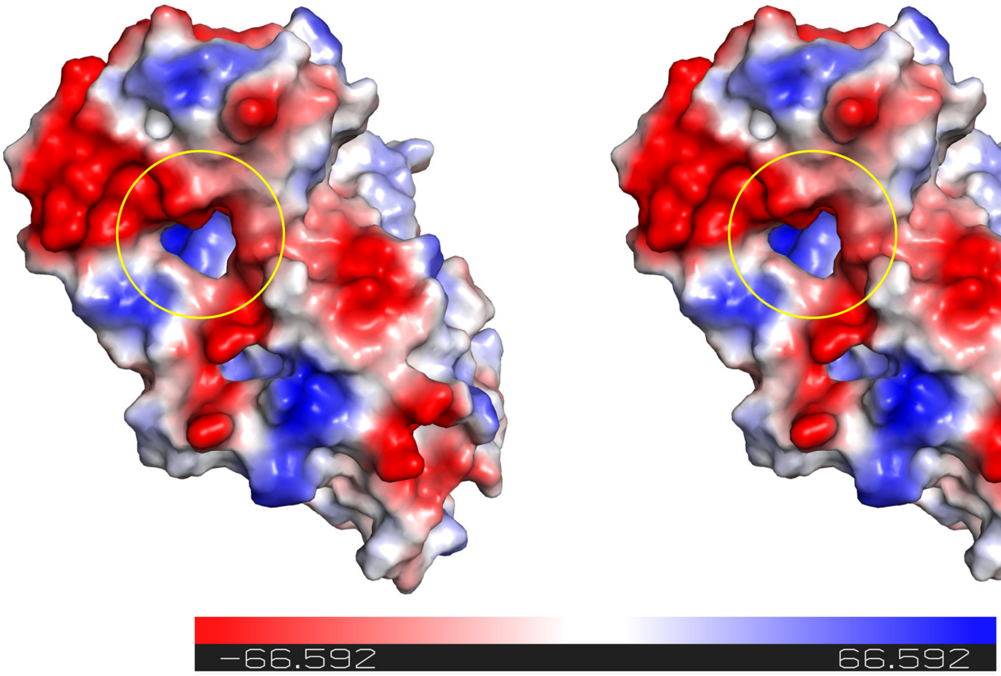


Fig 4-43: Stereo view of electrostatic potentials mapped on the surface of AmbD in complex with α-KG. Red areas indicate negative and blue positive potential. Substrate binding cavity and possible catalytic area is marked with a yellow circle, in which two residues Arg^{178} and Arg^{270} contribute a highly positively charged area.

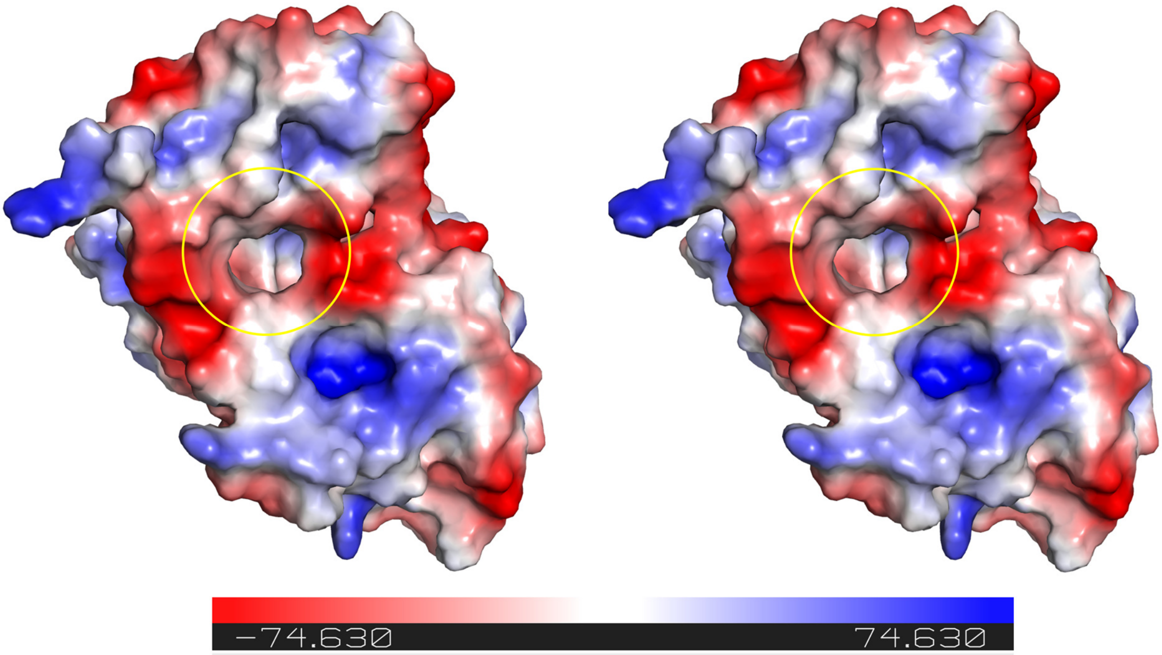


Fig 4-44: Stereo view of electrostatic potentials mapped on the surface of AmbC 30-350 in complex with NOG. Arg^{155} contributes the positively charged area near the active cavity, which is marked with a yellow circle.

In AmbC, in contrast, only one positively charged Arg^{155} residue was found near the active site, resulting with less possible attraction to the *L*-Glu moiety (Fig 4-44). As predicted, both AmbC and AmbD performed hydroxylation on *L*-Glu moiety, which would possibly happen on β and γ carbon sites, respectively. While the timing of these tailoring hydroxylation reactions is unclear, the difference between numbers of Arginine residue may influence the substrate binding specificity and timing of reactions.

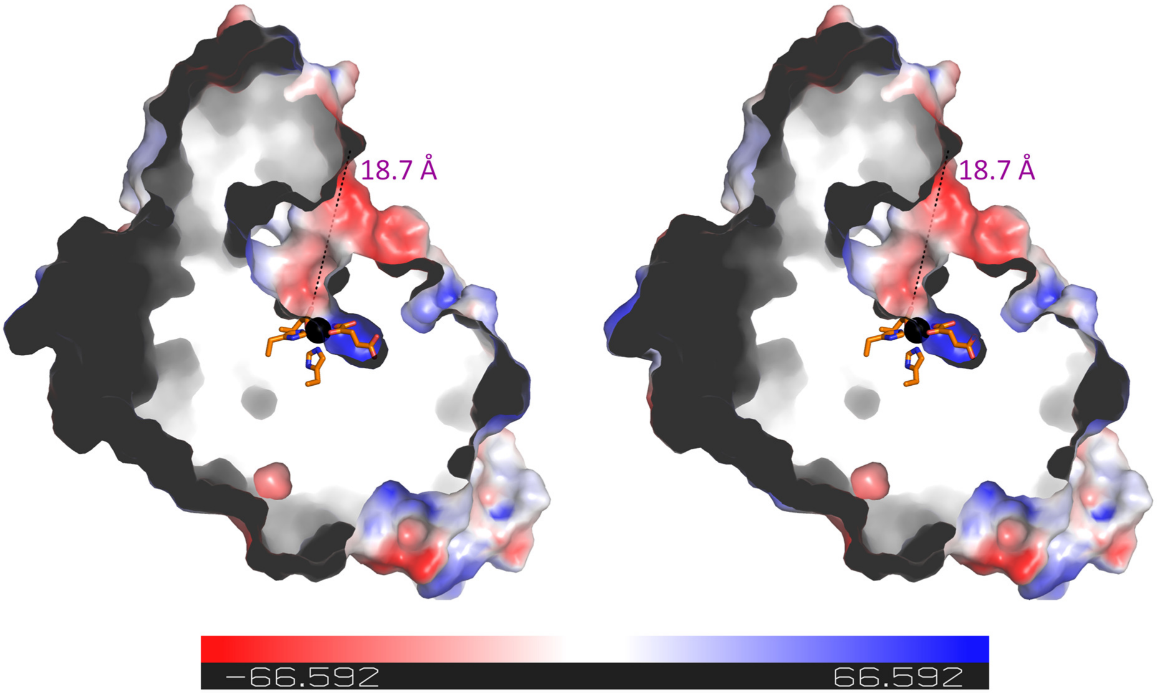


Fig 4-45: Stereo view of electrostatic potentials mapped on the surface of AmbD in complex with α-KG from the cross section view. Conserved residues and co-substrate α-KG are shown as sticks colored by elements (carbon atoms are colored in orange). The metal cobalt is shown as sphere colored in black. Distance was measured with PyMOL.

A cross section of the complex structure of AmbD with α-KG was generated to show the depth of the cavity. The distance from metal binding site to cavity edge is around 18.7 Å (Fig 4-45). As AMB is synthesized under a thiotemplate mechanism, the substrates or intermediates will be tethered to a phosphopantetheine arm obtained from Coenzyme A (Elovson & Vagelos, 1968) (Fig 4-46). This phosphopantetheine arm was measured around 17 Å, together with the length of potential dipeptide *L*-Ala-*L*-Glu or tripeptide *L*-

Ala-*L*-Glu-*L*-Ala, the L-Glu moiety can be well catched by the active site of the complex structure of AmbD with α-KG. As for the complex structure of AmbC 30-350 with NOG, the depth of its cavity was measured around 18.6 Å (Fig 4-47), which is almost the same as in the complex structure of AmbD with α-KG. Based on this structural analysis, these two enzymes could therefore perform potential hydroxylations to the substrate or intermediate near the active site cavity.

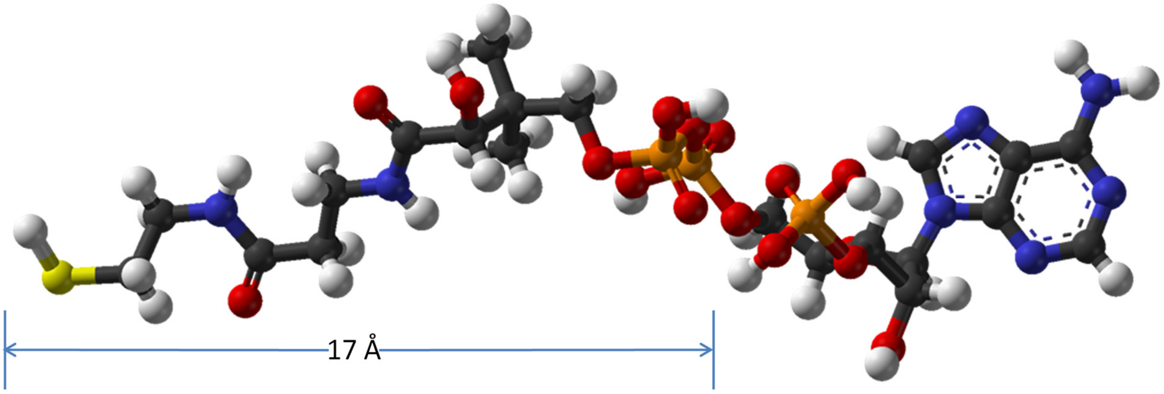


Fig 4-46: 3D structure of coenzyme A. The length of phosphopantetheine arm was measured at 17 Å by MolView (T. J. Smith, 1995).

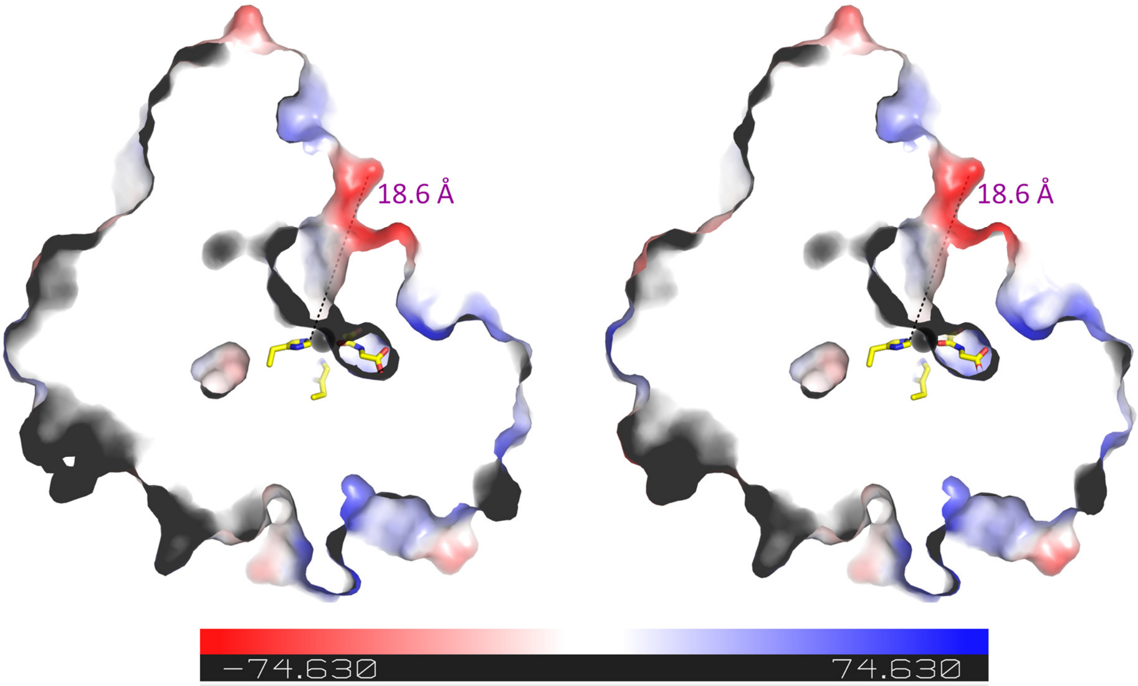


Fig 4-47: Stereo view of electrostatic potentials mapped on the surface of AmbC 30-350 complex with NOG from the cross section view. Conserved residues and co-substrate analog NOG are shown as sticks colored by elements (carbon atoms are colored in yellow). The metal nickel is shown as sphere colored in black. Distance was measured with PyMOL.

The electrostatic potential surface was set to show as cavities and pockets to check the potential reaction channel. One large pocket with a 22.2 Å width was observed near the active site (Fig 4-48). Under the same cavity detection conditions, a smaller pocket with a 16.6 Å width was found in the complex structure of AmbC 30-350 with NOG (Fig 4-49).

Considering the specificity to substrate or intermediate, the size of pocket may influence the sequence of tailoring reactions in AMB biosynthesis. As initial substrates for AMB synthesis are two *L*-Ala and one *L*-Glu, these three building blocks will be condensed and modified under the functions of AmbB, AmbE, AmbC and AmbD to produce final tripeptide analog *L*-Ala-AMB-*L*-Ala, which is then released by the thioesterase domain of AmbE. So the hydroxylation of *L*-Glu moiety performed by AmbC and AmbD may happen to *L*-Glu alone or dipeptide *L*-Ala-*L*-Glu or tripeptide *L*-Ala-*L*-Glu-*L*-Ala. Taking the size of pocket into account, the hydroxylation performed by AmbD may happen after AmbC.

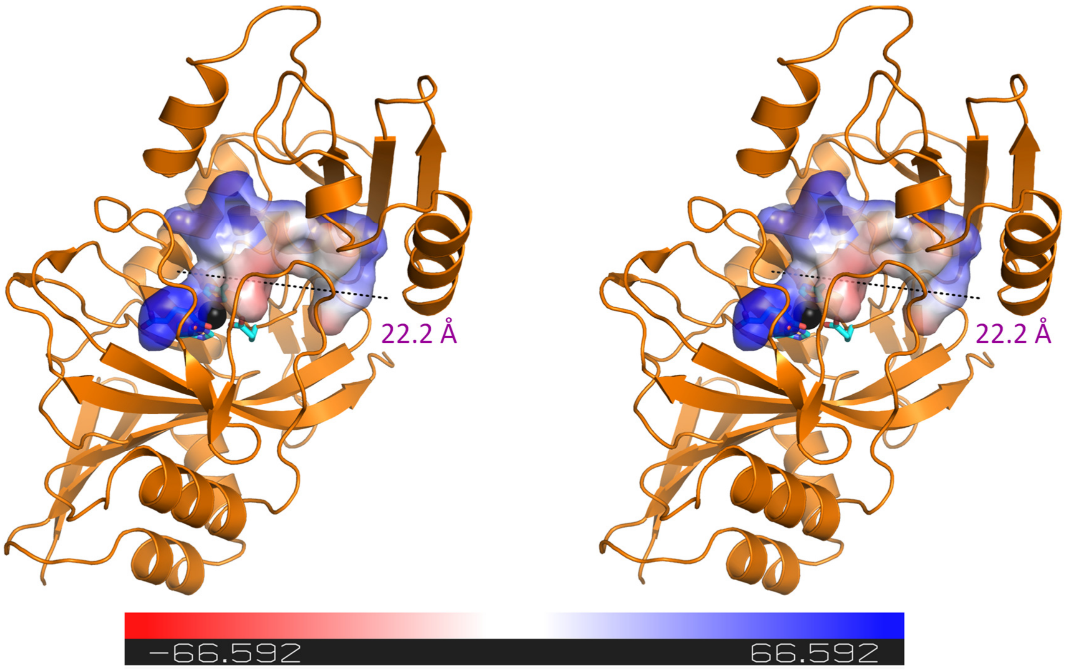


Fig 4-48: Stereo view of electrostatic potential surface shown as cavity and pocket mapped on the complex structure of AmbD with α-KG. The cavity detection radius and cavity detection cutoff were both set at 4 solvent radii. Conserved residues and co-substrate α-KG are shown as sticks colored by elements (carbon atoms are colored in cyan). The metal cobalt is shown as sphere colored in black. Distance was measured with PyMOL.

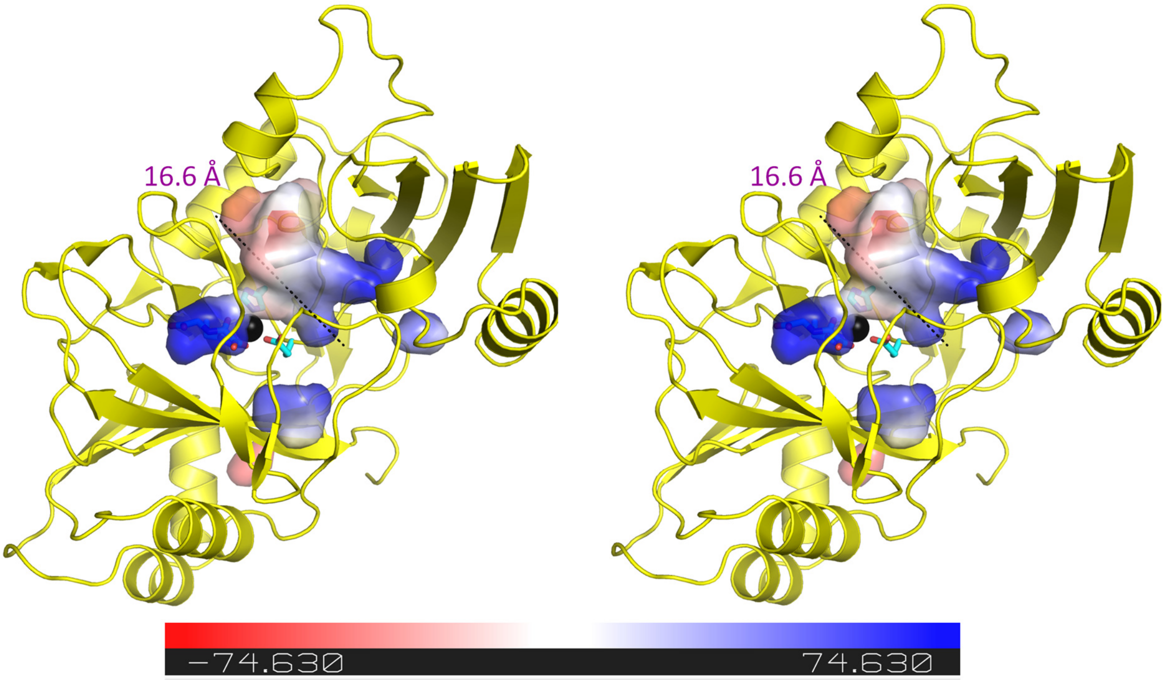


Fig 4-49: Stereo view of electrostatic potential surface shown as cavity and pocket mapped on the complex structure of AmbC with NOG. The cavity detection radius and cavity detection cutoff were both set at 4 solvent radii. Total four cavities were defined, while three smaller ones were disconnected with the substrate binding one. Conserved residues and co-substrate NOG are shown as sticks colored by elements (carbon atoms are colored in cyan). The metal nickel is shown as sphere colored in black. Distance was measured with PyMOL.

4.4 Comparison between AmbC/D and related proteins

4.4.1 Structural comparison with homologous structures

The Dali server is a network service for comparing protein structures in 3D (Holm & Laakso, 2016; Holm & Rosenstrom, 2010). By comparing with other structures from the Protein Data Bank (PDB), clues to reveal how AmbC and AmbD bind with their substrates may be obtained. The structure of AmbC 30-350 in complex with NOG was submitted to the Dali server. Several hits were reported, and the top 5 hits are shown in table 4-17.

Table 4-17: the top 5 homologous hits to AmbC 30-350 in complexwith NOG obtained from DALI server, and all of them are oxygenase. The color codes of the proteins reflect the colors applied in the following superposition.

Protein	PDB-ID	PyMOL r.m.s.d. (Å)	DALI r.m.s.d. (Å)	Sequence identity (%)
AmbD	-	1.825	-	49
At3g21360	1Y0Z	2.43	2.1	29
ScoE	6DCH	5.705	3.7	15
CarC	4OJ8	5.006	2.8	15
dioxygenase from *M. abscessus*	4Y0E	5.371	2.9	18
PvcB (PA2255)	3EAT	5.345	2.8	26

Except the gene product from *Arabidopsos thaliana* At3g21360 (PDB: 1Y0Z (Bitto *et al.*, 2005)), being taken as the search model for molecular replacement, the other four hits possess higher r.m.s.d. values to AmbC 30-350 in complex with NOG. They are the isonitrile biosynthesis enzyme ScoE from *Streptomyces coeruleorubidus* (PDB: 6DCH (Harris *et al.*, 2018)), the carbapenem synthase CarC from *Pectobacterium carotovorum* (PDB: 4OJ8 (Chang *et al.*, 2014)), the putative dioxygenase from *Mycobacterium abscessus* (released PDB: 4Y0E, to be published) and the dioxygenase PvcB protein from *Pseudomonas aeruginosa* (PDB: 3EAT (Drake & Gulick, 2008)).

With low sequence identity and high r.m.s.d. values, the structural superposition between AmbC 30-350 in complex with NOG/AmbD in complex with AKG and the top 5 hits show some varieties (Fig 4-50A). However, the conserved “jelly roll” sheets could be observed (Fig 4-50B).

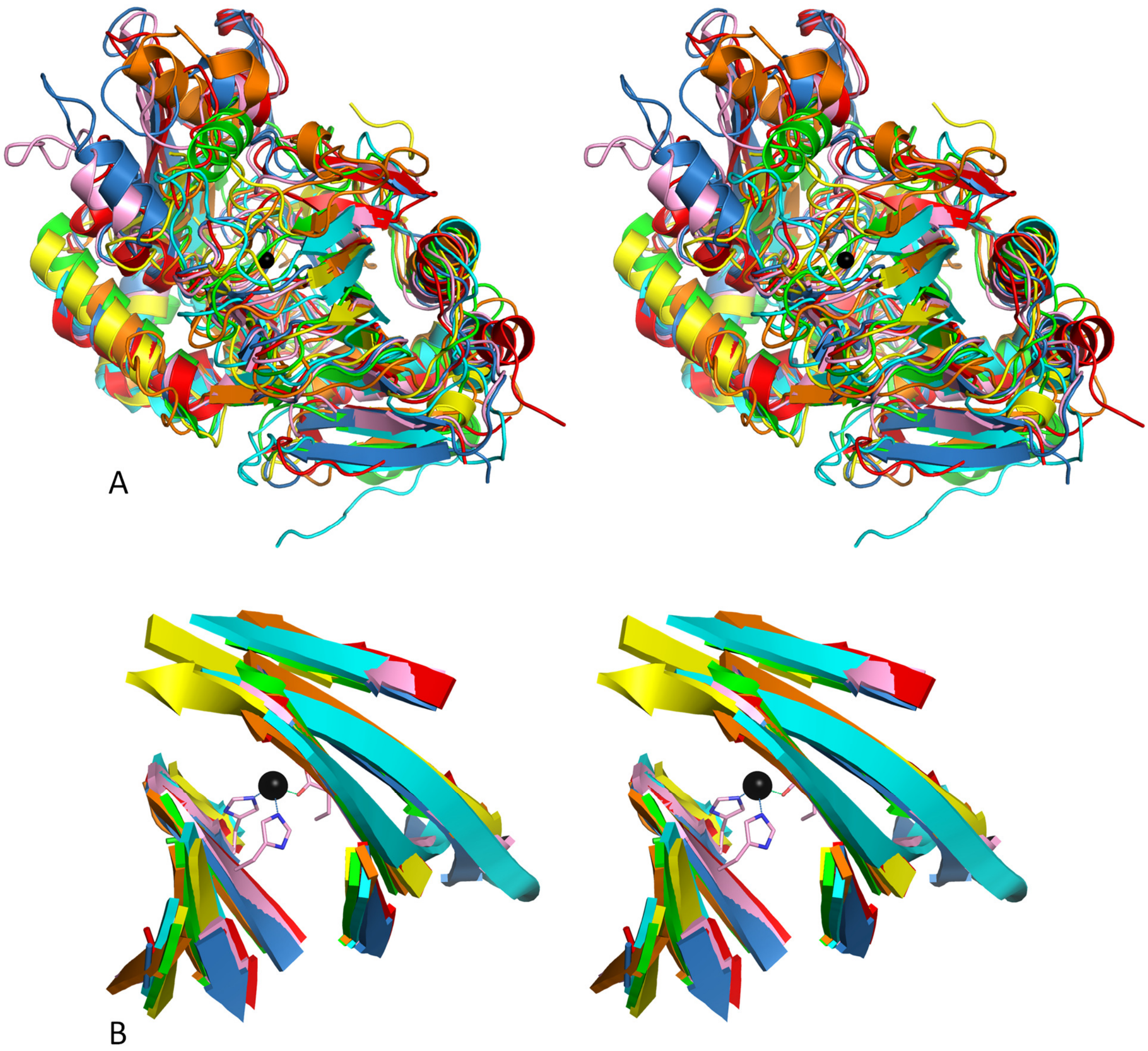


Fig 4-50: Stereo view of the overall structure and jelly roll sheets superposition. A: The structures of AmbC 30-350 in complex with NOG and AmbD in complex with AKG were both applied to present the superposition. All structures are shown as cartoon. Pink: AmbC 30-350 in complex with NOG, Blue: AmbD in complex with AKG, Red: At3g21360, Orange: ScoE, Yellow: CarC, Green: dioxygenase from *M. abscessus*, Cyan: PvcB. The peripheral areas are slightly variable. For orientation, the central metal of AmbC 30-350 in complex with NOG is shown as sphere colored black. B: Superposition of the conserved

jelly roll sheets. Color information is the same as in A. The metal coordinated residues from AmbC 30-350 in complex with NOG are shown as sticks colored pink.

4.4.2 AmbC and AmbD as the α-ketoglutarate-dependent dioxygenases

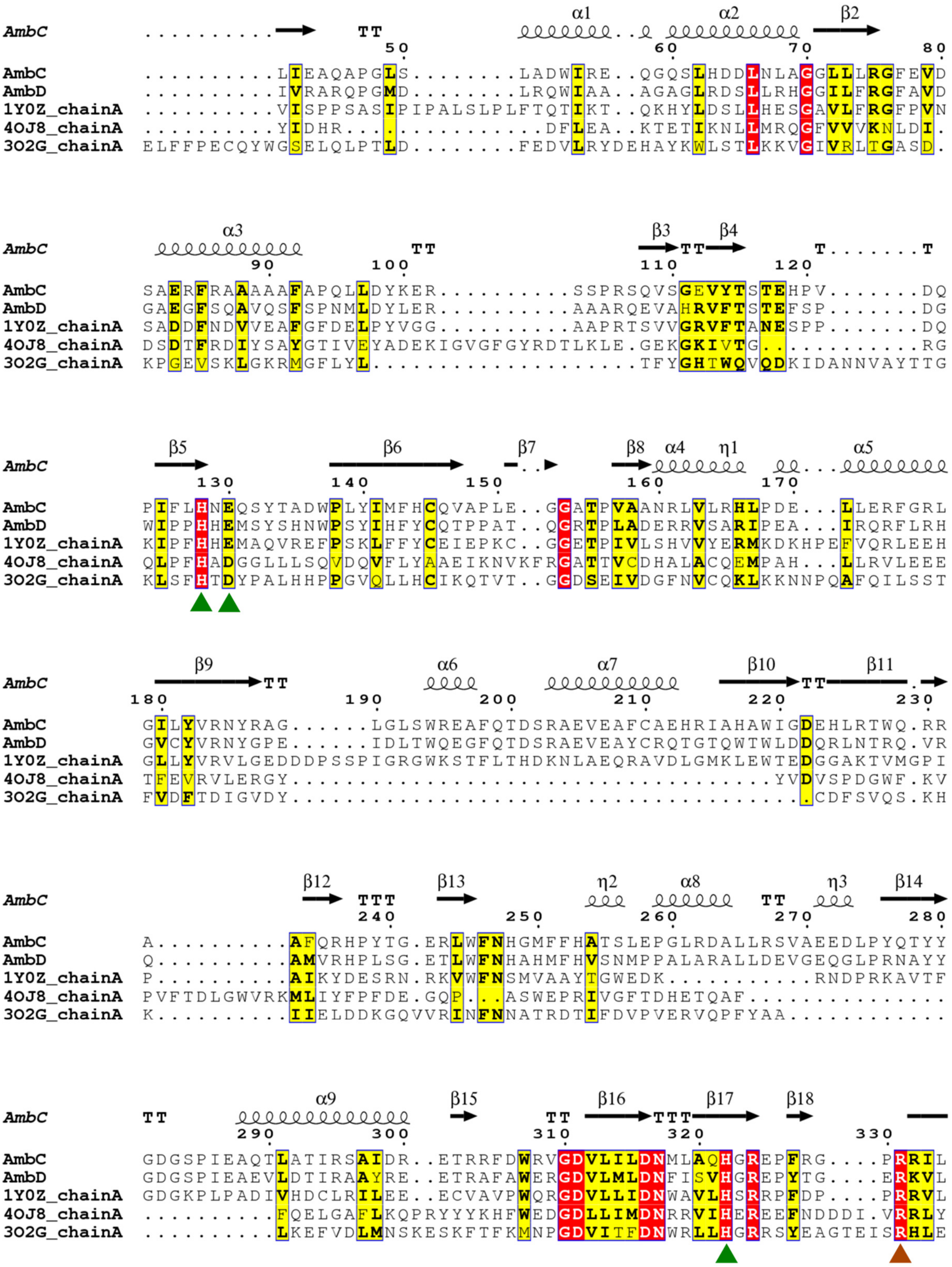
AmbC and AmbD are predicted as α-ketoglutarate-dependent dioxygenases. Two enzymes of this family of which structures have been deposited in the PDB were chosen for comparison with AmbC 30-350 in complex with NOG and AmbD in complex with AKG. These two deposited structures contain carboxylic substrates with characteristics similar to the *L*-Glu moiety. One is the carbapenem synthase (CarC) in complex with (3S, 5S)-carbapenam (Fig 4-51A, code as 2TQ) (PDB: 4OJ8 (Chang *et al.*, 2014)), and the other one is the human gamma-butyrobetaine 2-oxoglutarate dioxygenase 1 (BBOX1) in complex with 3-carboxy-N, N, N-trimethyl propan-1-aminium (Fig 4-51B, code as NM2) (PDB: 3O2G, to be published).

(3S,5S)-carbapenam (code: 2TQ)

3-carboxy-N,N,N-trimethylpropan-1-aminium (code: NM2)

Fig 4-51: Structural formula of (3S, 5S)-carbapenam (A) and 3-carboxy-N,N,N-trimethyl propan-1-aminium (B). Both chemicals contain a carboxyl group, which is similar to the *L*-Glu moiety.

Together with the highest homologous protein At3g21360 (PDB: 1Y0Z (Bitto *et al.*, 2005)) to AmbC, all five protein sequences were aligned. It seems that conserved residues mostly build the β-strands that compose the core jelly roll β-sheets. Three conserved metal-coordinating residues (two histidines and one carboxylic acid) and one co-substrate (NOG or AKG) coordinated arginine were observed (Fig 4-52).



```
AmbC        ..........LIEAQAPGLS........LADWIRE..QGQSLHDDLNLAGGLLLRGFEVD
AmbD        ..........IVRARQPGMD........LRQWIAA..AGAGLRDSLLRHGGILFRGFAVD
1Y0Z_chainA ..........VISPPSASIPIPALSLPLFTQTIKT..QKHYLDSLLHESGAVLFRGFPVN
4OJ8_chainA ..........YIDHR...............DFLEA..KTETIKNLLMRQGFVVVKNLDI.
3O2G_chainA ELFFPECQYWGSELQLPTLD........FEDVLRYDEHAYKWLSTLKKVGIVRLTGASD.

AmbC        SAERFRAAAAAFAPQLLDYKER..........SSPRSQVSGEVYTSTEHPV.......DQ
AmbD        GAEGFSQAVQSFSPNMLDYLER..........AAARQEVAHRVFTSTEFSP.......DG
1Y0Z_chainA SADDFNDVVEAFGFDELPYVGG..........AAPRTSVVGRVFTANESPP.......DQ
4OJ8_chainA DSDTFRDIYSAYGTIVEYADEKIGVGFGYRDTLKLE.GEKGKIVTG............RG
3O2G_chainA KPGEVSKLGKRMGFLYL....................TFYGHTWQVQDKIDANNVAYTTG

AmbC        PIFLHNEQSYTADWPLYIMFHCQVAPLE..GGATPVAANRLVLRHLPDE...LLERFGRL
AmbD        WIPPHHEMSYSHNWPSYIHFYCQTPPAT..QGRTPLADERRVSARIPEA...IRQRFLRH
1Y0Z_chainA KIPFHHEMAQVREFPSKLFFYCEIEPKC..GGETPIVLSHVVYERMKDKHPEFVQRLEEH
4OJ8_chainA QLPFHADGGLLLSQVDQVFLYAAEIKNVKFRGATTVCDHALACQEMPAH...LLRVLEEE
3O2G_chainA KLSFHTDYPALHHPPGVQLLHCIKQTVT..GGDSEIVDGFNVCQKLKKNNPQAFQILSST

AmbC        GILYVRNYRAG......LGLSWREAFQTDSRAEVEAFCAEHRIAHAWIGDEHLRTWQ.RR
AmbD        GVCYVRNYGPE......IDLTWQEGFQTDSRAEVEAYCRQTGTQWTWLDDQRLNTRQ.VR
1Y0Z_chainA GLLYVRVLGEDDDPSSPIGRGWKSTFLTHDKNLAEQRAVDLGMKLEWTEDGGAKTVMGPI
4OJ8_chainA TFEVRVLERGY....................................YVDVSPDGWF.KV
3O2G_chainA FVDFTDIGVDY.......................................CDFSVQS.KH

AmbC        A..........AFQRHPYTG.ERLWFNHGMFFHATSLEPGLRDALLRSVAEEDLPYQTYY
AmbD        Q..........AMVRHPLSG.ETLWFNHAHMFHVSNMPPALARALLDEVGEQGLPRNAYY
1Y0Z_chainA P..........AIKYDESRN.RKVWFNSMVAAYTGWEDK...........RNDPRKAVTF
4OJ8_chainA PVFTDLGWVRKMLIYFPFDE.GQP...ASWEPRIVGFTDHETQAF...............
3O2G_chainA K..........IIELDDKGQVVRINFNNATRDTIFDVPVERVQPFYAA............

AmbC        GDGSPIEAQTLATIRSAIDR..ETRRFDWRVGDVLILDNMLAQHGREPFRG....PRRIL
AmbD        GDGSPIEAEVLDTIRAAYRE..ETRAFAWERGDVLMLDNFISVHGREPYTG....ERKVL
1Y0Z_chainA GDGKPLPADIVHDCLRILEE..ECVAVPWQRGDVLLIDNWAVLHSRRPFDP....PRRVL
4OJ8_chainA ..........FQELGAFLKQPRYYYKHFWEDGDLLIMDNRRVIHEREEFNDDDI.VRRLY
3O2G_chainA ..........LKEFVDLMNSKESKFTFKMNPGDVITFDNWRLLHGRRSYEAGTEISRHLE
```

Fig 4-52: Protein sequence alignment of AmbC with AmbD, At3g21360 (PDB: 1Y0Z), CarC (PDB: 4OJ8) and BBOX1 (PDB: 3O2G). This figure was generated with ESPript (Robert & Gouet, 2014) and adjusted manually to optimize the structure-based alignment.

Secondary structure elements and turns (TT) of AmbC 30-350 in complex with NOG are shown. Identical residues are boxed and colored in red. Majority of conserved residues build the β-strands. Metal-coordinating residues (two histidines and one carboxylic acid) are marked with a green triangle below the alignment, and one arginine coordinated with co-substrate (NOG or AKG) an orange triangle.

4.4.3 The arrangement of the active site

The superposition of the active site residues of these four structures (AmbC 30-350 in complex with NOG, AmbD in complex with AKG, CarC and BBOX1) showed high similarity. The co-substrate AKG in the structure of CarC is replaced with NOG in the structure of BBOX1. The highly conserved metal-coordinating residues (two histidines and one carboxylic acid) were observed. For the structures of AmbC 30-350 in complex with NOG and AmbD in complex with AKG, two arginine and two threonine residues were identified to coordinate with respective co-substrate. One extra arginine contacting with the C_5-carboxylate oxygen of AKG/NOG through a hydrogen bond can be found in the structures of CarC and BBOX1, while the latter protein does not provide the arginine and threonine that bind the C_1-carboxylate and C_5-carboxylate oxygens, respectively. In the case of CarC, the less conserved arginine is located a bit further away from the central metal, resulting in forming one hydrogen bond with the C_1-carboxyl group of AKG (Fig 4-53).

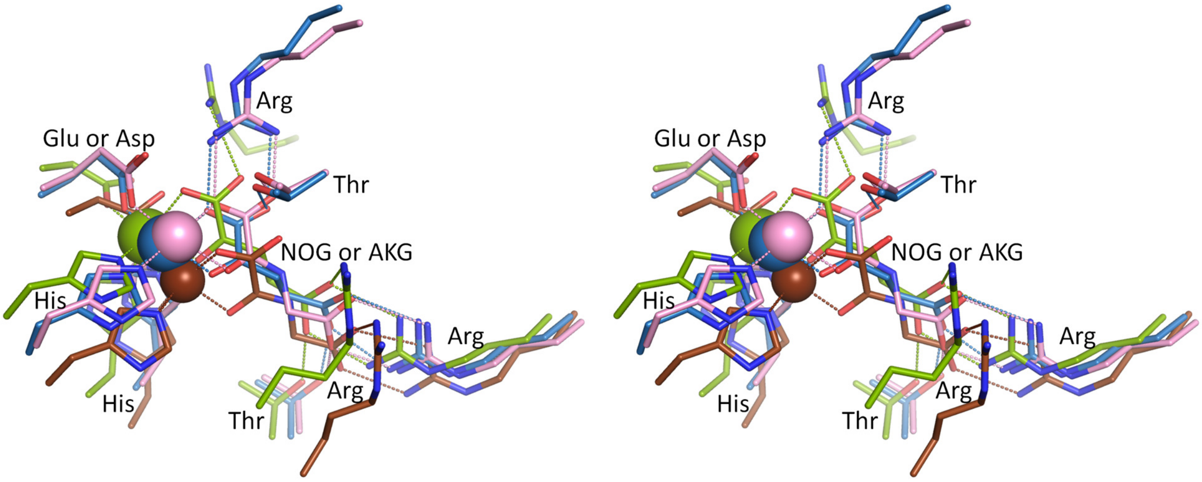


Fig 4-53: The stereo view of the active site superposition of AmbC 30-350 in complex with NOG (pink), AmbD in complex with NOG (blue), CarC (green) and BBOX1 (brown). The metal-coordinating residues show high similarity in all four structures. For co-substrate-coordinating residues, AmbC 30-350 in complex with NOG and AmbD in complex with

AKG also share the similar configuration. One unusual hydrogen bond formed between the C5-carboxylate oxygen of AKG/NOG and the extra arginine is typical for CarC and BBOX1.

In the structure of CarC, the specific substrate (3S, 5S)-carbapenam (2TQ) lies above the metal site and binds with the residues Tyr^{164}, Arg^{267} and Gln^{269}. All the coordination is conducted through the carboxyl group of 2TQ (Fig 4-54A). Similar to CarC, in the structure of BBOX1, the substrate 3-carboxy-N, N, N-trimethyl propan-1-aminium (NM2) also locates above the metal site and coordinates to residues Asn^{191} and Asn^{292} through the hydrogen bonds formed by carboxyl group of NM2 (Fig 4-54B).

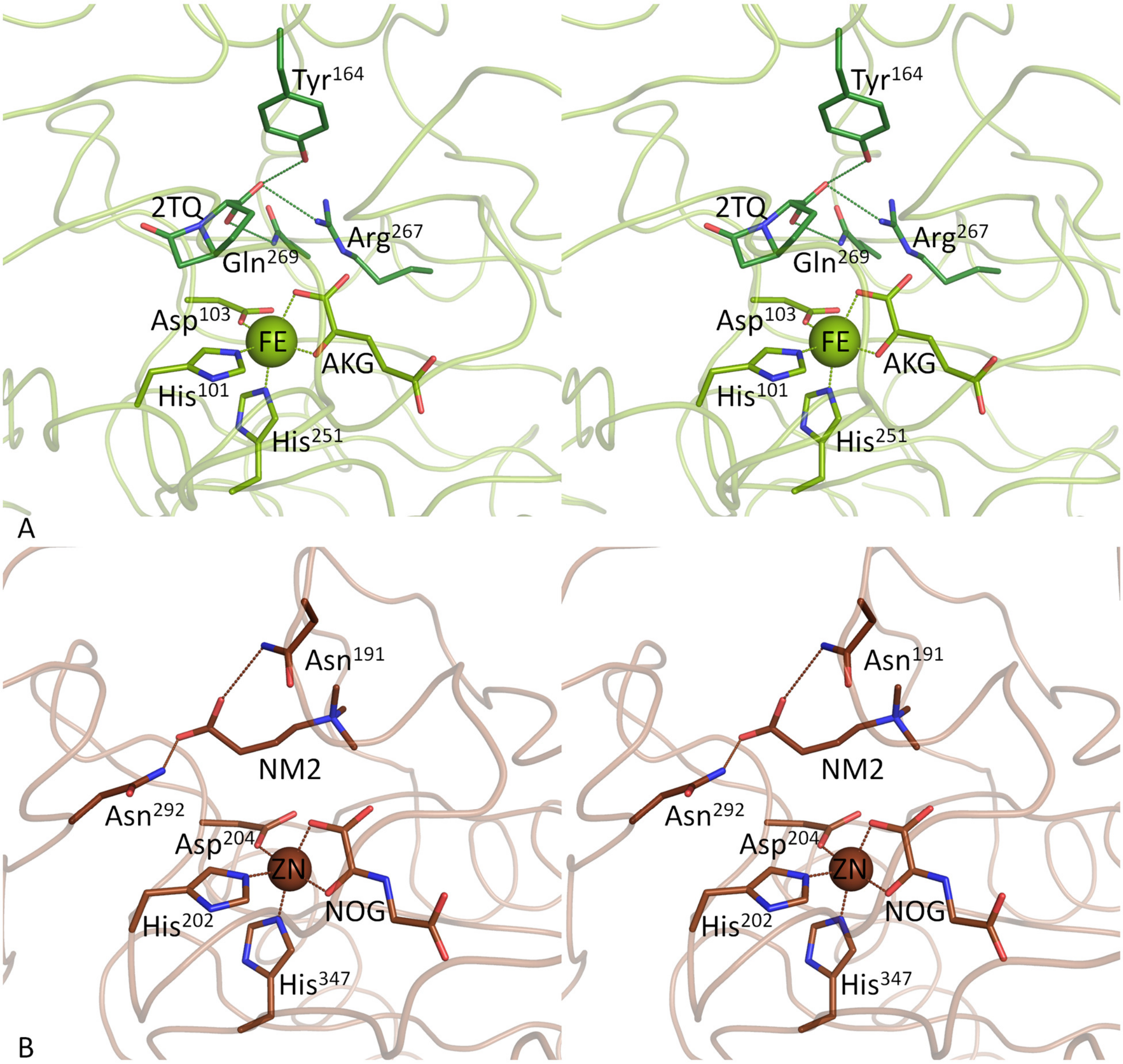


Fig 4-54: Stereo view of active sites of CarC (green) and BBOX1 (brown). A: In the structure of CarC, the substrate 2TQ contacts with residues Tyr^{164}, Arg^{267} and Gln^{269}. The central iron is shown as sphere labeled FE of which the coordinated residues and co-substrate AKG are labeled. B: The substrate NM2 contacts with Asn^{191} and Asn^{292} in the case of BBOX1. The central zinc is shown as sphere labeled ZN, and the coordinating residues and the ligand NOG are labeled.

Subsequently, a superposition of AmbC 30-350 in complex with NOG and AmbD in complex with AKG was performed. Based on the position of 2TQ and NM2 in the CarC and BBOX1 structures, several potential binding residues surrounding them in the superposition were identified. In the structure of AmbC 30-350 in complex with NOG, the residues Arg^{73}, Arg^{155}, His^{219} and Gln^{102} are found nearby the active site. Accordingly, all residues (Arg^{96}, Arg^{178} and His^{242}) can be traced in the structure of AmbD in complex with AKG except the Gln^{102}, which is typical for AmbC. Among them, Arg^{155}/Arg^{178} locates on a β-strand, and the Arg^{73}/Arg^{96} locates on a flexible loop (see details in the B-factor analysis). The latter arginine residue stretches further away from the active site, while the flexibility of the loop may still make the binding with specific substrate possible (Fig 4-55).

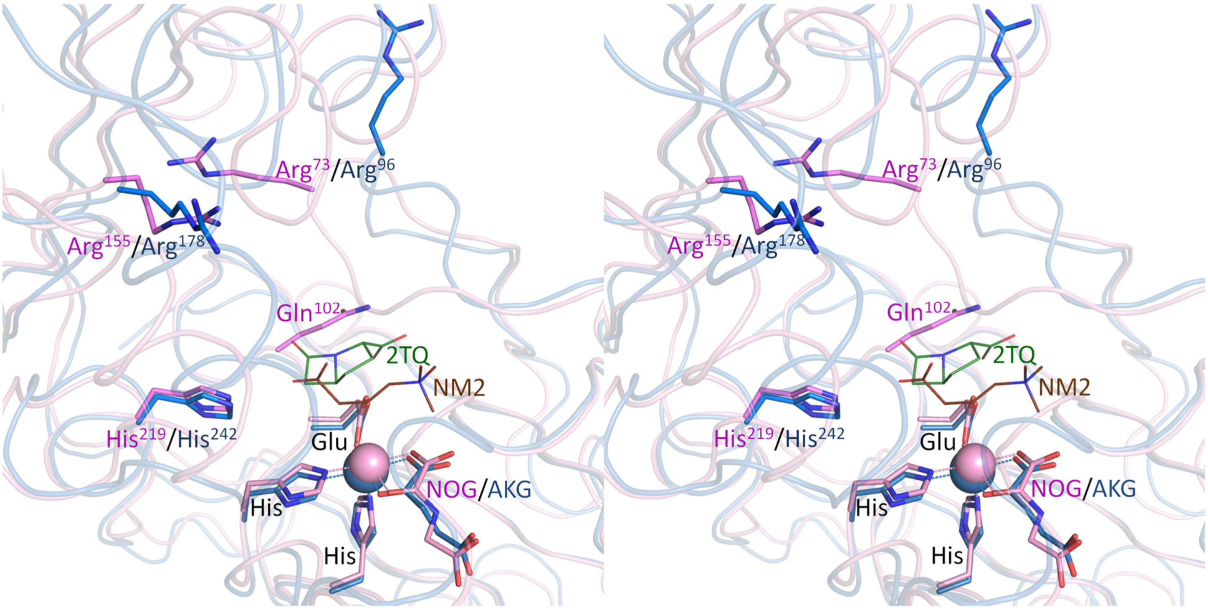


Fig 4-55: Stereo view of the superposition of the active sites of AmbC 30-350 in complex with NOG (pink) and AmbD in complex with AKG (blue). For orientation, the substrates 2TQ (green) and NM2 (brown) from CarC and BBOX1, respectively, are shown as lines. The potential binding residues to specific substrate Arg^{73}/Arg^{96}, Arg^{155}/Arg^{178}, His^{219}/His^{242}

and the typical Gln^{102} for AmbC are shown. Metal-coordinating residues and ligands are also shown, and the overall structures are shown as ribbon.

4.5 The structure of uncharacterized domain of AmbE

4.5.1 Identification and analysis of unknown domain AmbE 20-420

As mentioned in the introduction, AmbE is a Non-Ribosomal Peptide Synthetase. It is a multi-domain protein and contains 2124 amino acids. Its full-length sequence was submitted to InterPro (protein sequence analysis and classification) online server to analyze different domains and classify them according to their functions (Hunter *et al.*, 2009). The result is shown in Fig 4-56. There are six domains in total, which are an amino acid adenylation domain (A), an S-adenosyl-*L*-methionine-dependent methyltransferase domain (MT), a phosphopantetheine binding ACP domain (T1), a condensation domain (C), another phosphopantetheine binding ACP domain (T2) and a thioesterase domain (TE), respectively.

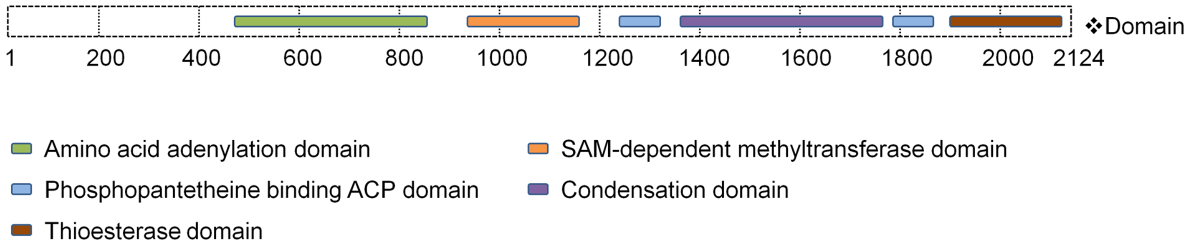


Fig 4-56: AmbE sequence analysis and domains classification. Six domains in total could be defined and five functional domains were classified.

Interestingly, no prediction could be made for the first 450 amino acids, which should form one domain at least. Sequence analysis results of AmbE1-450 from PSIPRED and FoldIndex online server (Buchan *et al.*, 2013; Jaime Prilusky, 2005) are shown in Fig 4-57 and Fig 4-58, respectively. The first 20 amino acids and the last 25 amino acids possess a high possibility of disorder and are therefore likely unfolded, which will decrease the tendency to form crystals. Therefore, the truncated sequence AmbE 20-440 was cloned, overexpressed and purified, but no crystal could be found after plenty of trials. Subsequently, another truncated sequence AmbE 20-420 was chosen. This sequence was furthermore submitted to the structure prediction server Phyre2 (Kelley *et al.*, 2015) and five top hits focused on synthetase, albeit with low sequence identity, were reported (Table 4-18).

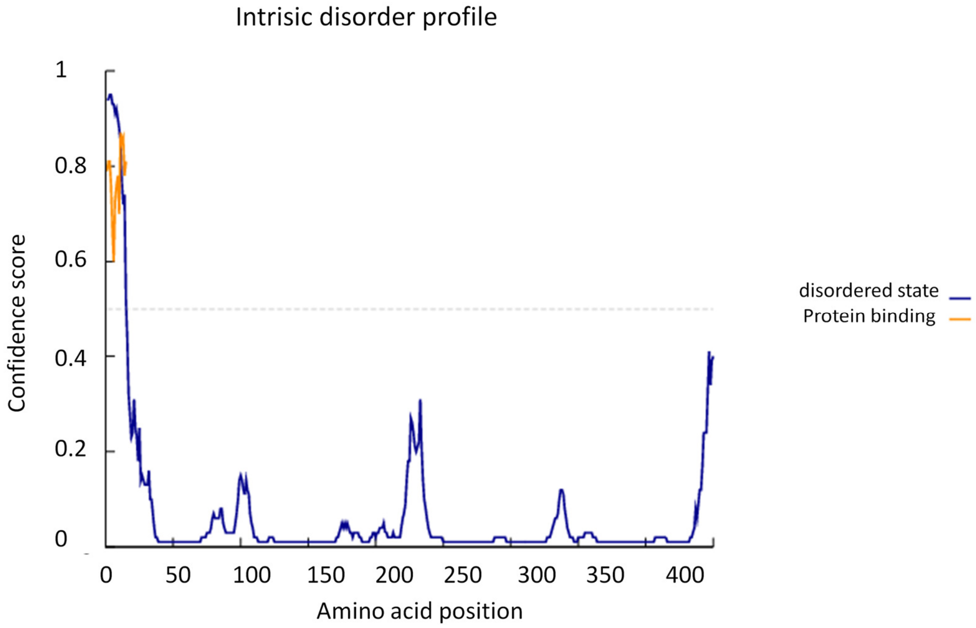


Fig 4-57: DISOPRED analysis of AmbE 1-450. The blue line represents the likelihood of disorder. Amino acids located at the N terminus were predicted as disordered, because the blue line is above the grey dashed line, which indicates a confidence score of 0.5. The orange line shows the confidence of disordered protein binding residue predictions. This sequence is probably the COM domain, which can selectively interact with other NRPS proteins (Hahn & Stachelhaus, 2004).

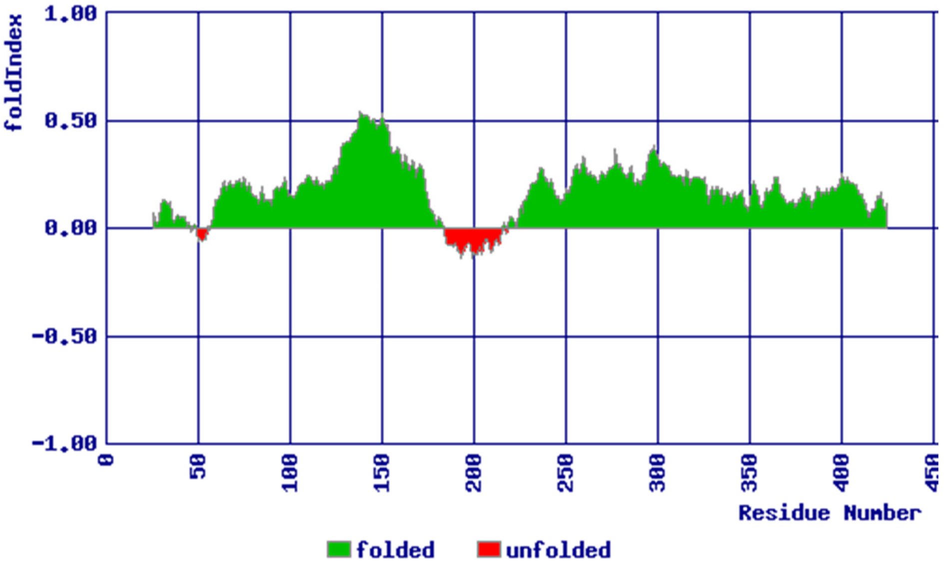


Fig 4-58: FoldIndex of AmbE 1-450. The missing green signal for the N-terminal 20 amino acids and the C-terminal 30 amino acids indicates that these regions are flexible. The red unfolded sequences represent predicted flexible loops.

Table 4-18

Top 5 templates used for structure prediction of AmbE 20-420.

PDB ID	Function annotation	Sequence identity / %	Model confidence /%
4ZXI	tyrocidine synthetase 3;	13	100
2VSG	Ligase (surfactin synthetase subunit 3)	13	100
2JGP	Ligase (tyrocidine synthetase 3)	15	100
4ZNM	Ligase (c-domain type ii peptide synthetase)	17	100
4ZXJ	enterobactin synthase component f;	16	100

4.5.2 Purification and crystallization of AmbE 20-420

The coding sequence of residues AmbE 20-420 was PCR amplified by using the pET27b *ambE* plasmid as the template. The PCR product was inserted between the KpnI and HindIII restriction sites of pOPINB. The protein AmbE 20-420 was expressed and purified as usual. The final size-exclusion chromatography (SEC) purification results were identified by SDS-PAGE (Fig 4-59)

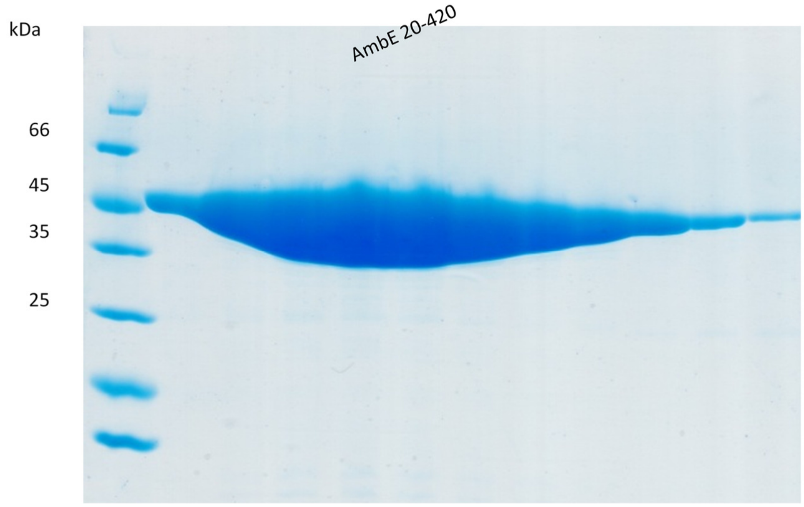


Fig 4-59: SDS-PAGE followed by InstantBlue staining for protein AmbE 20-420 after SEC purification.

Initial crystallization condition screens were set up for AmbE 20-420. Two hits were identified (JCSG III E3 and Morpheus G2), and both conditions yielded high-quality crystals (Fig 4-60 and 61). After in-house testing, seven crystals were sent to beamline P11 in DESY (Burkhardt, 2016). 3600 images of 0.1° rotation were collected on a Pilatus 6M detector at a wavelength of λ = 1.0332 Å. The best dataset was analyzed and detailed statistics are shown in Table 4-19.

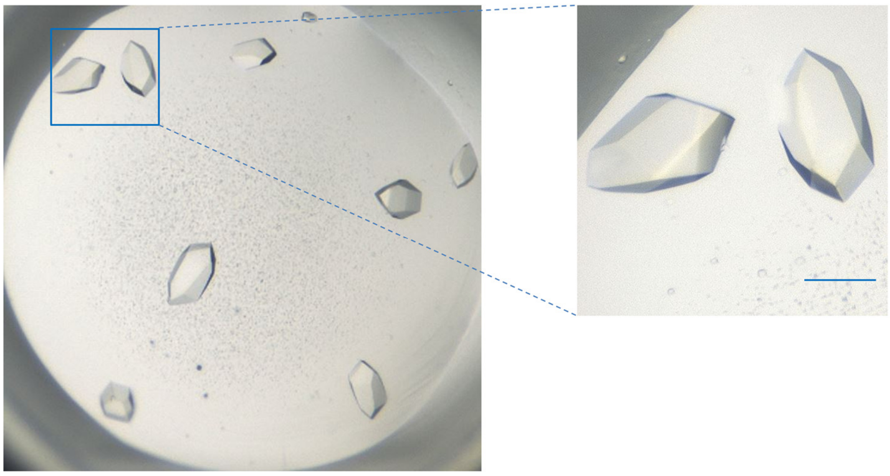


Fig 4-60: Crystals of AmbE 20-420 obtained from JCSG III E3 (0.1 M MES pH 6.5, 1.6 M ammonium sulfate and 10% 1, 4-dioxane). Scale bar is 100 µm.

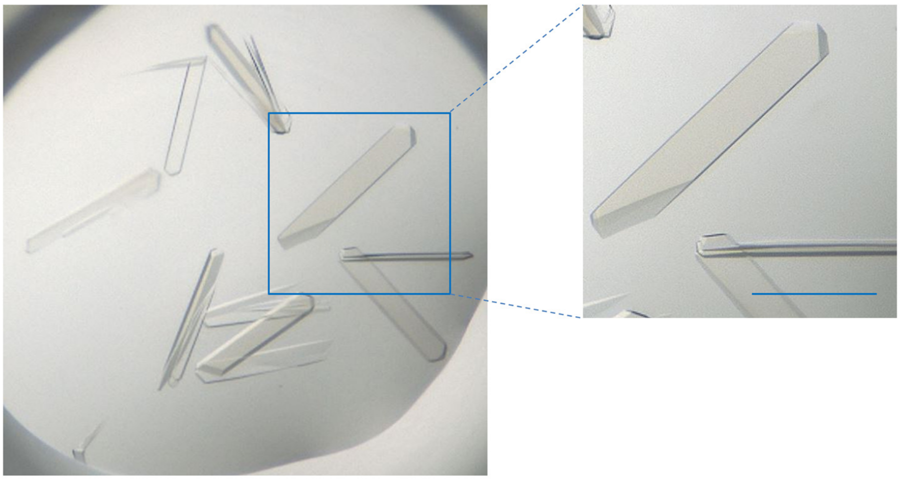


Fig 4-61: Crystals of AmbE 20-420 obtained from Morpheus G2 (0.1 M Morphous buffer 1 pH 6.5, 0.1 M Morpheus carboxylic acid and 30% EDO_P8K). Scale bar is 100 µm.

Table 4-19: Data collection statistics for AmbE 20-420 generated with the aimless program (Evans, 2011; Evans PR, 2013). Values in parentheses are for the highest resolution shell.

Dataset	AmbE 20-420
Wavelength (Å) / beamline	1.0332 / DESY, P11
Space group	C1 2 1
Resolution range (Å)	46.03-2.5 (2.6-2.5)
Unit cell parameters (°)	90, 103.02, 90
(Å)	93.99, 92.06, 56.26
Mosaicity (°)*	0.12
Rpim	3.5 (43.3)
Total number of observations	111262 (12550)
Total number unique	16174 (1806)
Mean I / σ (I)	14.6 (2)
Mn (I) half-set correlation CC (1/2) (%)	99.9 (82)
Completeness (%)	99.7 (99.5)
Multiplicity	6.9 (6.9)

*Mosaicity value is from XDS (Kabsch, 2010)

The crystal of AmbE 20-420 diffracted to a resolution of 2.5 Å at P11 beamline in DESY and reflections were indexed in space group C121 with cell dimensions a = 93.99, b = 92.06, c = 56.26. There wes one molecule in the asymmetric unit according to Matthews_coef calculation in CCP4 (Kantardjieff & Rupp, 2003; Matthews, 1968). No solution could be found by molecular replacement methods with this dataset.

Selono-*L*-methionine (SeMet) labeled protein AmbE 20-420 was purified, and initial crystallization screens were set up. Suitable crystals were obtained under JCSG IV D10 and Morpheus E2 (Fig 4-62 and 63). Three crystals were sent to beamline X06DA-PXIII in SLS (Waltersperger *et al.*, 2015), and six datasets with three different chi angles were collected. 3600 images of 0.1° rotation were collected on a Pilatus 2M-F detector at a wavelength of λ = 0.979390 Å. The best dataset was analyzed and detailed statistics are shown in Table 4-19.

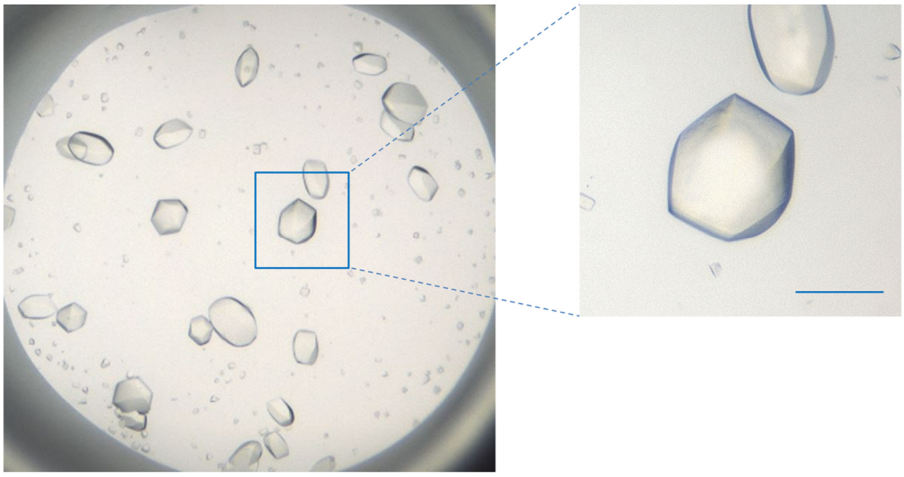


Fig 4-62: Crystals of SeMet AmbE 20-420 obtained from JCSG IV D10 (0.085 M HEPES pH7.5, 1.7% PEG400, 15% glycerol and 1.7 M ammonium sulfate). Scale bar is 100 µm.

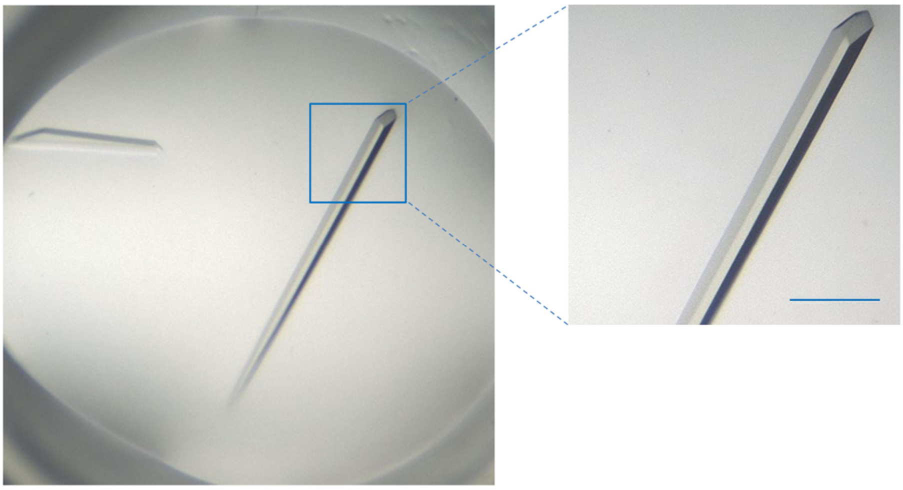


Fig 4-63: Crystals of SeMet AmbE 20-420 obtained from Morpheus E2 (0.1 M Morpheus buffer 1, 0.12 M Morpheus ethylene glycols and 30% EDO_P8K). Scale bar is 100 µm.

Table 4-20: Data collection statistics for SeMet AmbE 20-420 genereated with aimless program (Evans, 2011; Evans PR, 2013). Values in parentheses are for the highest resolution shell.

Dataset	SeMet AmbE 20-420
Wavelength (Å) / beamline	0.979390 / SLS, PXIII
Space group	$P3_1$ 2 1
Resolution range (Å)	45.74-2.1 (2.16-2.10)
Unit cell parameters (°)	90, 90, 120

(Å)	90.22, 90.22, 112.83
Mosaicity (°)*	0.09
Rpim	2.0 (39.1)
Total number of observations	1266103 (95459)
Total number unique	31579 (2559)
Mean I / σ (I)	28.4 (2.2)
Mn (I) half-set correlation CC (1/2) (%)	100 (73.1)
Completeness (%)	100 (100)
Multiplicity	40.1 (37.3)

*Mosaicity value is from XDS (Kabsch, 2010)

The crystal of SeMet AmbE 20-420 diffracted to a resolution of 2.1 Å at the PXIII beamline of the SLS and reflections were indexed in space group $P3_121$ with cell dimensions a = 90.22, b = 90.22, c = 112.83. There was one molecule in the asymmetric unit according to Matthews_coef calculation in CCP4 (Kantardjieff & Rupp, 2003; Matthews, 1968).

4.5.3 Structure determination and refinement of SeMet and native AmbE 20-420

With the SeMet single-wavelength anomalous dispersion dataset, the AUTOSOL (Adams *et al.*, 2010; Terwilliger *et al.*, 2009) program from the PHENIX suite was employed to generate experimental phases and to build the initial structure of SeMet AmbE 20-420. This model was further applied to solve the structure of native AmbE 20-420. The phenix.refine program in Phenix (Headd *et al.*, 2014; Terwilliger, 2003) and manual adjustments performed in Coot (Emsley & Cowtan, 2004) followed to polish the final structure of SeMet AmbE 20-420 with R_{work} = 0.1980 and R_{free} = 0.2185 and 97.47% of the residues in Ramachandran-favored regions and a *MolProbity* (Chen *et al.*, 2010; Davis *et al.*, 2007) score of 1.22, as well as the native structure of Amb E20-420 with R_{work} = 0.2396 and R_{free} = 0.2690 and 95.58% of the residues in Ramachandran-favored regions and a *MolProbity* score of 1.67 (Table 4-21).

Table 4-21: Refinement statistics of SeMet and native AmbE 20-420 generated with Phenix (Diederichs & Karplus, 2013). Values in parentheses are for the highest resolution shell.

Dataset	SeMet AmbE 20-420	Native AmbE 20-420
Resolution range (Å)	45.74-2.1 (2.175-2.1)	39.83-2.5 (2.589-2.5)
Total reflections	1269299 (116700)	111662 (11115)
Unique reflections	31543 (3104)	16177 (1595)
Reflections used in refinement	31540 (3103)	16167 (1596)
Reflections used for R-free	1484 (152)	765 (83)
R_{work}	0.1980 (0.2961)	0.2396 (0.3603)
R_{free}	0.2185 (0.3017)	0.2690 (0.3633)
Protein residues	401	391
RMS (bonds)	0.004	0.004
RMS (angles)	0.95	0.93
Ramachandran favored (%)	97.47	95.58
Ramachandran outliers (%)	0	0.26
Rotamer outliers (%)	1.03	1.77
Clashscore	3.1	6.29
Average B-factor	53.74	76.63
macromolecules	53.9	76.58
ligands	60.13	88.26
solvent	44.78	71.42
MolProbity score*	1.22	1.67

*MolProbity value is from MolProbity (http://molprobity.biochem.duke.edu/) (Chen *et al.*, 2010; Davis *et al.*, 2007)

4.5.4 Overall structure of AmbE 20-420

The final refined SeMet AmbE 20-420 structure contains one molecule in the asymmetric unit. Around 20 amino acids located in N- and C-terminus are missing, and 2 gaps exist in between. For the native structure of AmbE 20-420, around 44 amino acids are missing. All residues that could be modeled in the two structures are shown in table 4-22. The two structures were very similar with an r.m.s.d. value of 1.705 Å (Fig 4-

64), and since the SeMet model was more complete due to higher resolution, it was chosen for further analysis.

Table 4-22: Residues of SeMet and native AmbE 20-420 structures that can be modeled at the resolution 2.1 Å and 2.5 Å respectively. For clarity, the first residue of the truncated AmbE 20-420 is numbered 1.

Chains	Start	End	Gaps
SeMet_Chain A	S12	P427	P172-A184; D226-G231
Native_Chain A	G13	R418	P172-A184; D226-G231

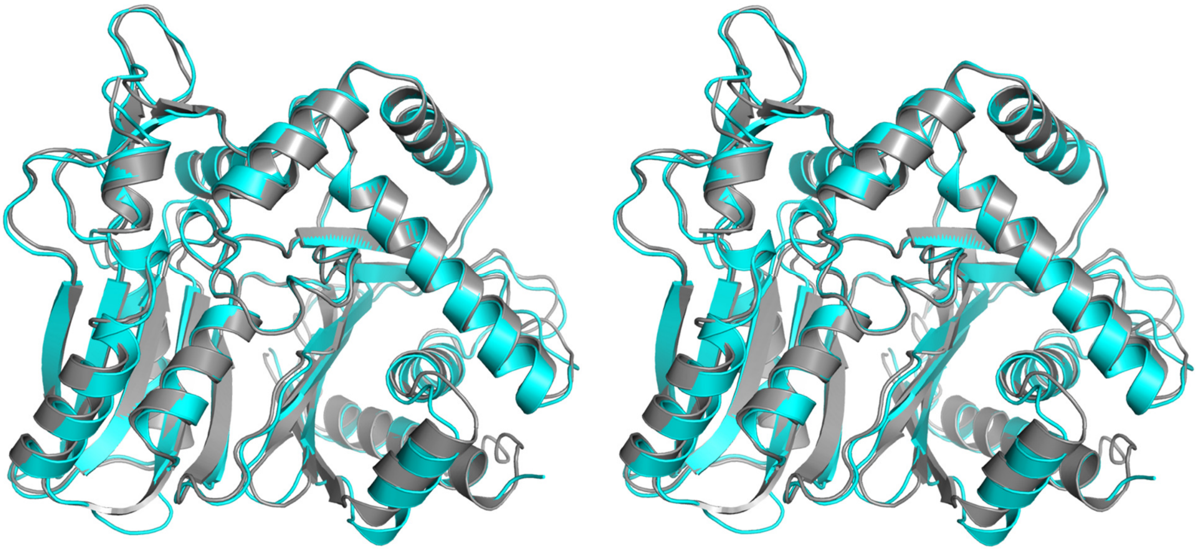


Fig 4-64: Stereo view of a superposition of SeMet (grey) and native AmbE 20-420 (cyan) structures.

4.5.5 Secondary and tertiary structure of AmbE 20-420

The structure of AmbE 20-420 was submitted to PDBsum (https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) (Laskowski *et al.*, 2018), which generated information about secondary structure and topology shown in Fig 4-65 and Fig 4-66. Each AmbE 20-420 monomer contained 3 β-sheets (14 strands) and 13 α-helixes.

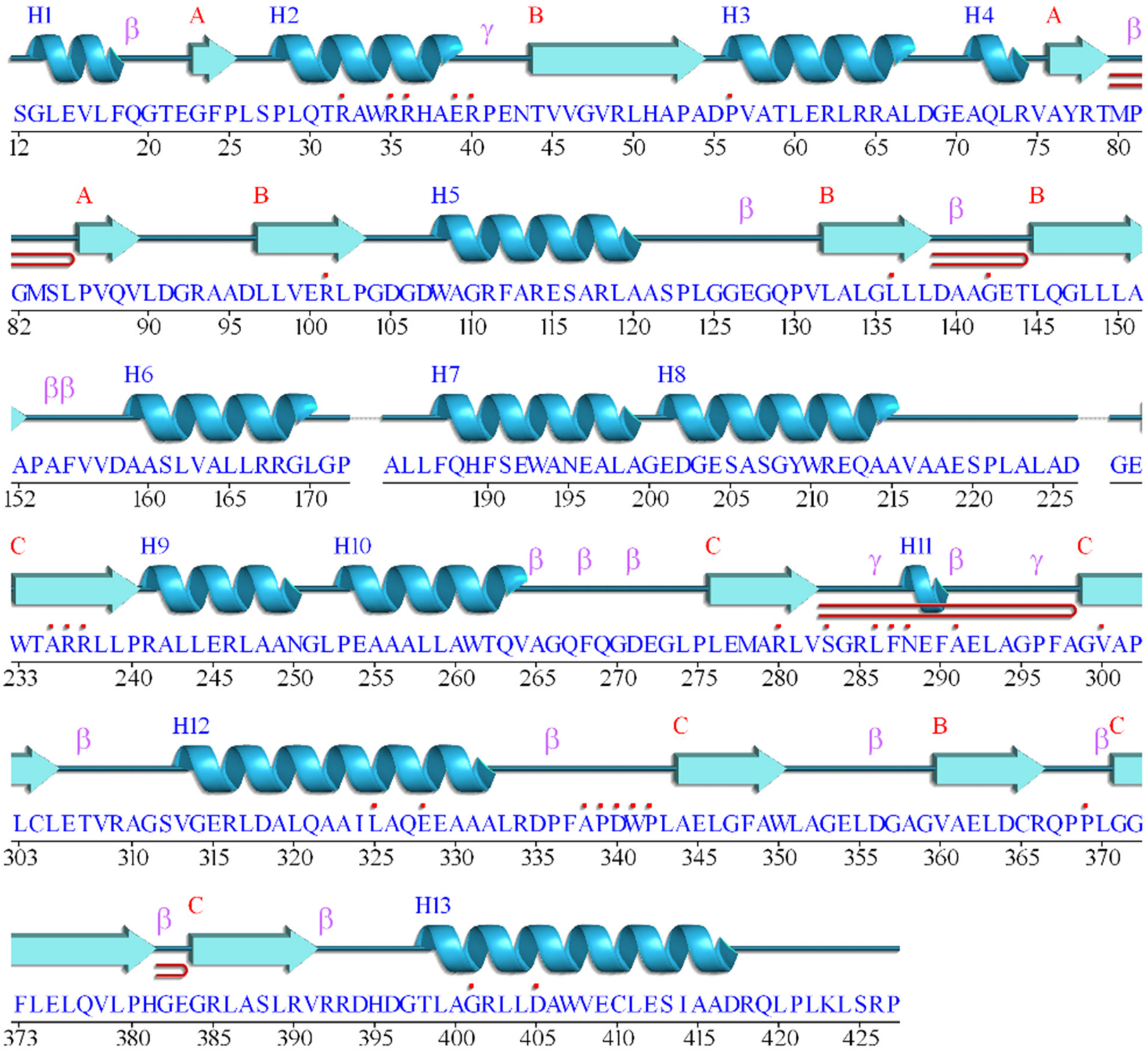


Fig 4-65: Primary and corresponding secondary structure of AmbE 20-420. The 13 α-helixes are shown as blue letters H1-H13, and 3 β-sheets formed by continuous β-strands are shown as red letters A-C. β means beta turn, γ means gamma turn, red clip means beta hairpin, red dot means residue contacts to ligands (polyethylene glycol molecules from the crystallization buffer).

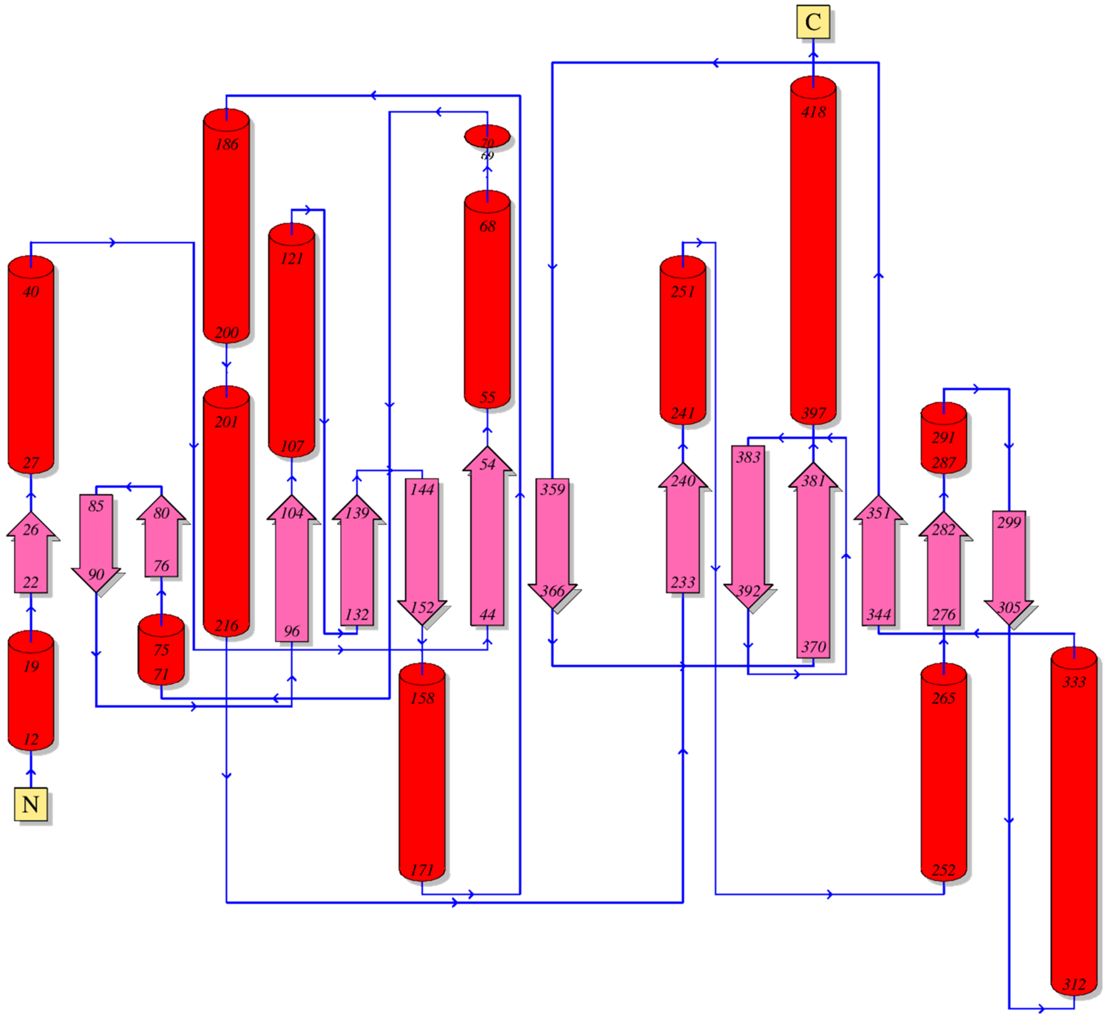


Fig 4-66: Topology of AmbE 20-420. Red columns represented 13 α-helixes, and pink arrows represented 14 β-strands classified to 3 β-sheets.

4.5.6 Structural analysis

4.5.6.1 B-factor analysis

The better defined SeMet AmbE 20-420 molecule contained 2 gaps and showed four loops with relatively high B-factors up to a maximum of 144 Å^2 (Fig 4-67). Among them, loop1 and loop2 displayed the highest B-factors. Together, they establish the biggest flexible area of AmbE 20-420. Loop3 and loop4 had B-factors around 100 Å^2, and the latter loop was far away from the other three. The central core area is relatively inflexible.

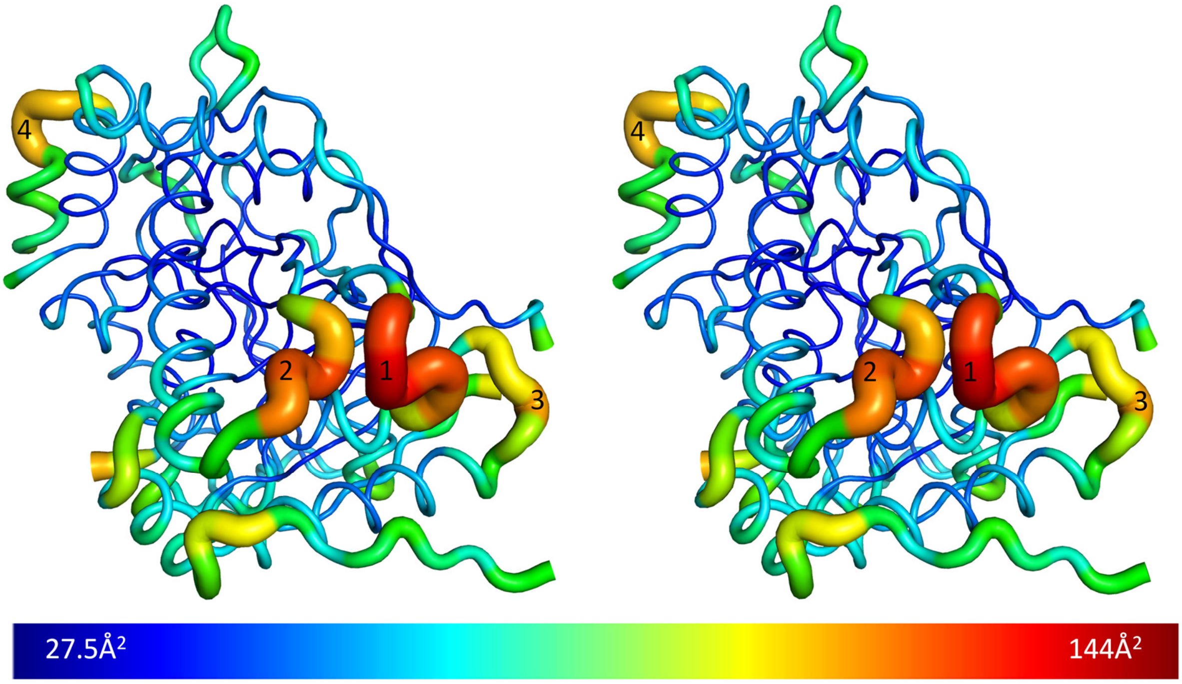


Fig 4-67: Stereo view on the B-factor distribution in the SeMet structure of AmbE 20-420. Wider ribbon and warmer colors represent higher B-factors. Four flexible loops were identified and labeled 1, 2, 3 and 4.

4.5.6.2 Surface analysis

The electrostatic potential calculated by the APBS plugin in PyMOL (Baker *et al.*, 2001) showed several positively and negatively charged patches on the molecular surface of AmbE 20-420. Especially for the sequence 192EWANEALAGEDGE204, forming a loop connecting two α-helixes, showed a highly negatively charged state (Fig 4-68A). A cross section view was generated from this orientation, revealing a channel that goes through the complete molecule (Fig 4-68B). The shape of the channel suggests that it may bind phosphopantetheine-linked molecules, similar torelated proteins (see below). Afterwards, the electrostatic potential surface was set to show as cavities and pockets. In addition to the major channel, several minor cavities located at the surface of the structure were identified (Fig 4-69).

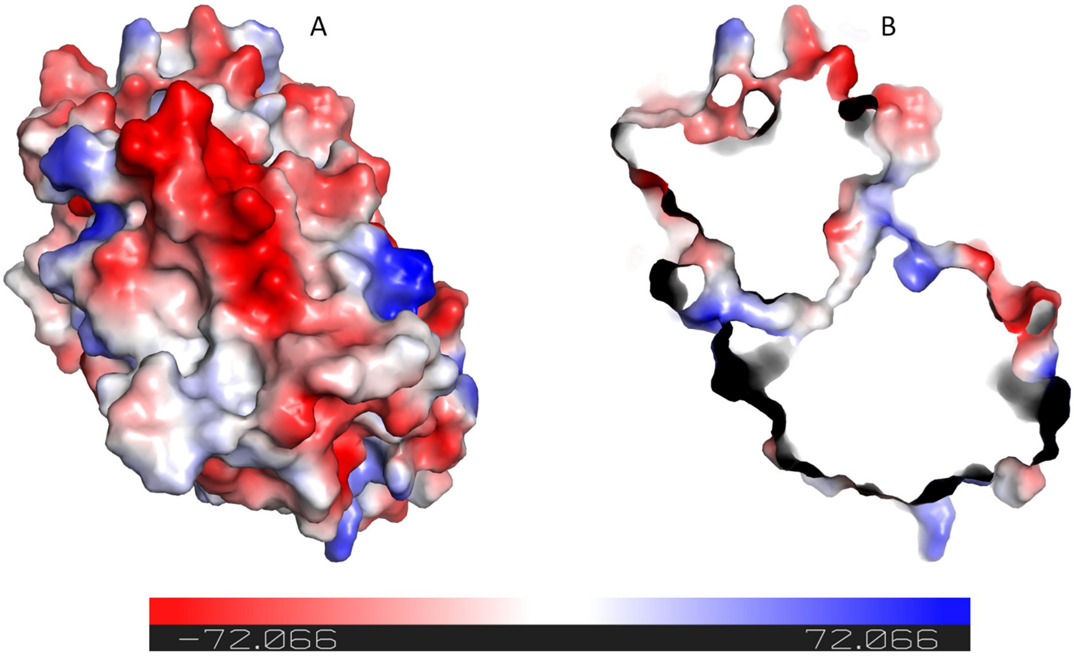


Fig 4-68: Electrostatic potentials mapped on the surface of AmbE 20-420. Red and blue areas indicate negative and positive potentials. A: One highly negatively charged area can be seen. B: One go-through tunnel is observed in the cross section view.

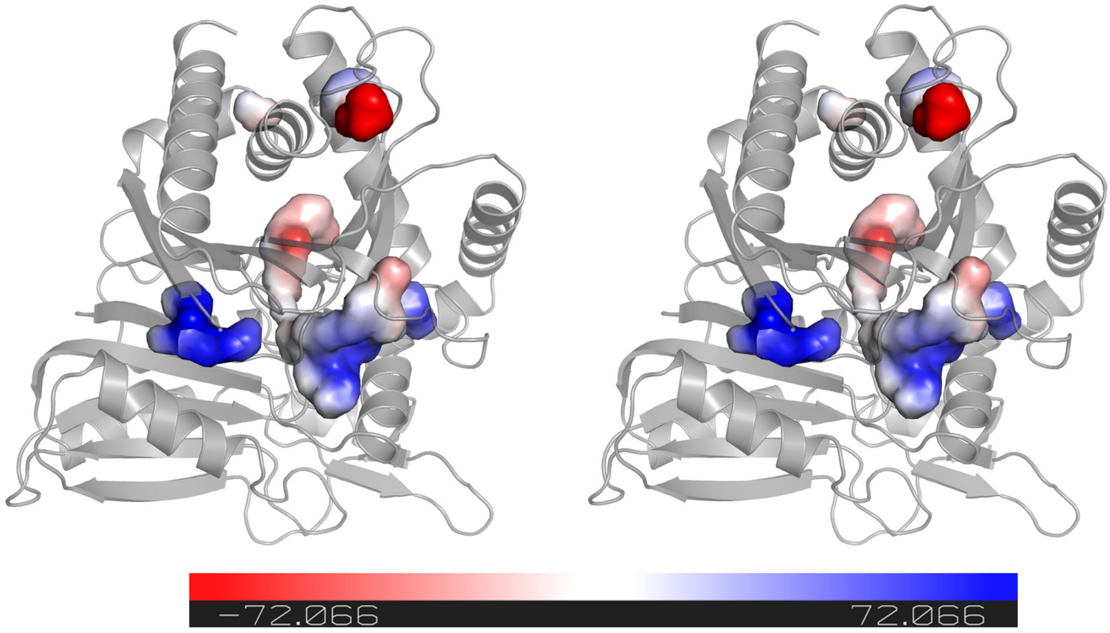


Fig 4-69: Stereo view of electrostatic potential surface shown as cavity and pocket mapped on the structure of AmbE 20-420. The cavity detection radius and cavity detection cutoff were set at 4 solvent radii and 3 solvent radii, respectively. The go-through tunnel was obviously defined and several highly charged cavities were also identified on the surface of the structure.

4.5.7 Function exploration

The Dali server is a network service for comparing protein structures in 3D (Holm & Laakso, 2016; Holm & Rosenstrom, 2010). By comparing with other structures from the Protein Data Bank (PDB), clues to reveal the function of uncharacterized proteins may be obtained. PDBeFold is another server to perform 3D alignment and predict the function of uncharacterized proteins (Henrick, 2005).

The structure of AmbE 20-420 was submitted to the Dali server and to PDBeFold respectively. The same top 5 hits were reported and shown in table 4-23. Three of them were condensation domains necessary for synthesizing peptides through a thiotemplate mechanism conducted by Non-Ribosomal Peptide synthetases (NRPSs). The other two were heterocyclization domains that introduce a ring into the final product. All these five hits shared very low sequence identity with AmbE 20-420, providing an explanation why molecular replacement with the native diffraction data was not successful. Superposition of AmbE 20-420 with the top hit CDA Peptide Synthetase I (PDB: 5DU9) (Bloudoff *et al.*, 2016) was performed, and the overall structures were similar with an r.m.s.d. value of 5.482 Å (Fig 4-70).

Table 4-23

Top 5 hits from structural alignment (DALI and PDBeFold)

PDB ID	Function annotation	Sequence identity / %	r.m.s.d. / Å
5DU9	CDA Peptide Synthetase I	16	3.9
4JN3	CDA Peptide Synthetase I	16	4.0
5T3E	Synthetase Heterocyclization Domain	14	3.8
4HVM	tallysomycin biosynthesis protein	18	3.7
5T81	cyclization-docking bidomain	16	3.7

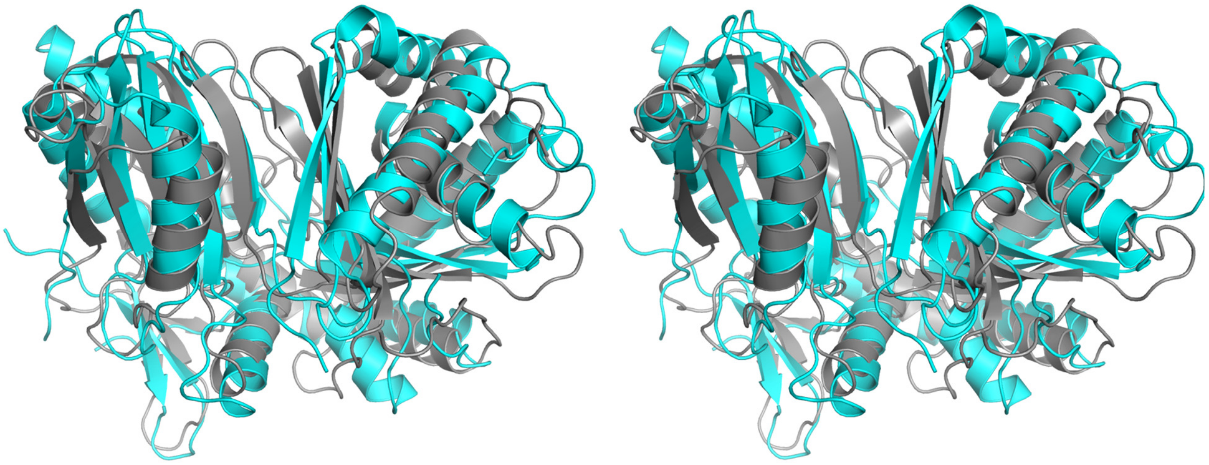


Fig 4-70: Stereo view of superposition between AmbE 20-420 and one condensation domain (PDB: 5DU9). AmbE 20-420 was colored in grey. 5DU9 was colored in cyan.

It has been reported that condensation domains contain a highly conserved HHXXXDG sequence motif, with the Asp and second His essential for activity (Bloudoff *et al.*, 2013; Keating *et al.*, 2002; Samel *et al.*, 2007), and its mutation abolishes the condensation reaction (Bergendahl *et al.*, 2002; Stachelhaus *et al.*, 1998).

Taking the sequence of 5DU9 as query template to run BLAST (Altschul *et al.*, 1990; Mount, 2007), another condensation domain (PDB: 1L5A) sequences was obtained (Keating *et al.*, 2002). Together with condensation domains of AmbB and AmbE, further alignment was performed and the conserved sequence motif HHXXXDG was observed (Fig 4-71). For the condensation domain of AmbB, the motif is changed slightly to HHXXXDD, which was reported earlier in another condensation domain MalF (Biggins *et al.*, 2012). A superposition of all available condensation structures is shown as well. As the most important residue, the second Histidine in this conserved HHXXXDG motif has been highlighted in Fig 4-72.

However, the conserved motif is absent in the sequence of AmbE 20-420, indicating that the uncharacterized sequence of AmbE cannot be a condensation domain. further, there are already two condensation domains within AmbB and AmbE respectively, which is enough to synthesize a tripeptide *L*-Ala-*L*-Glu-*L*-Ala, further corroborating that the N-terminal domain of AmbE must have a different function.

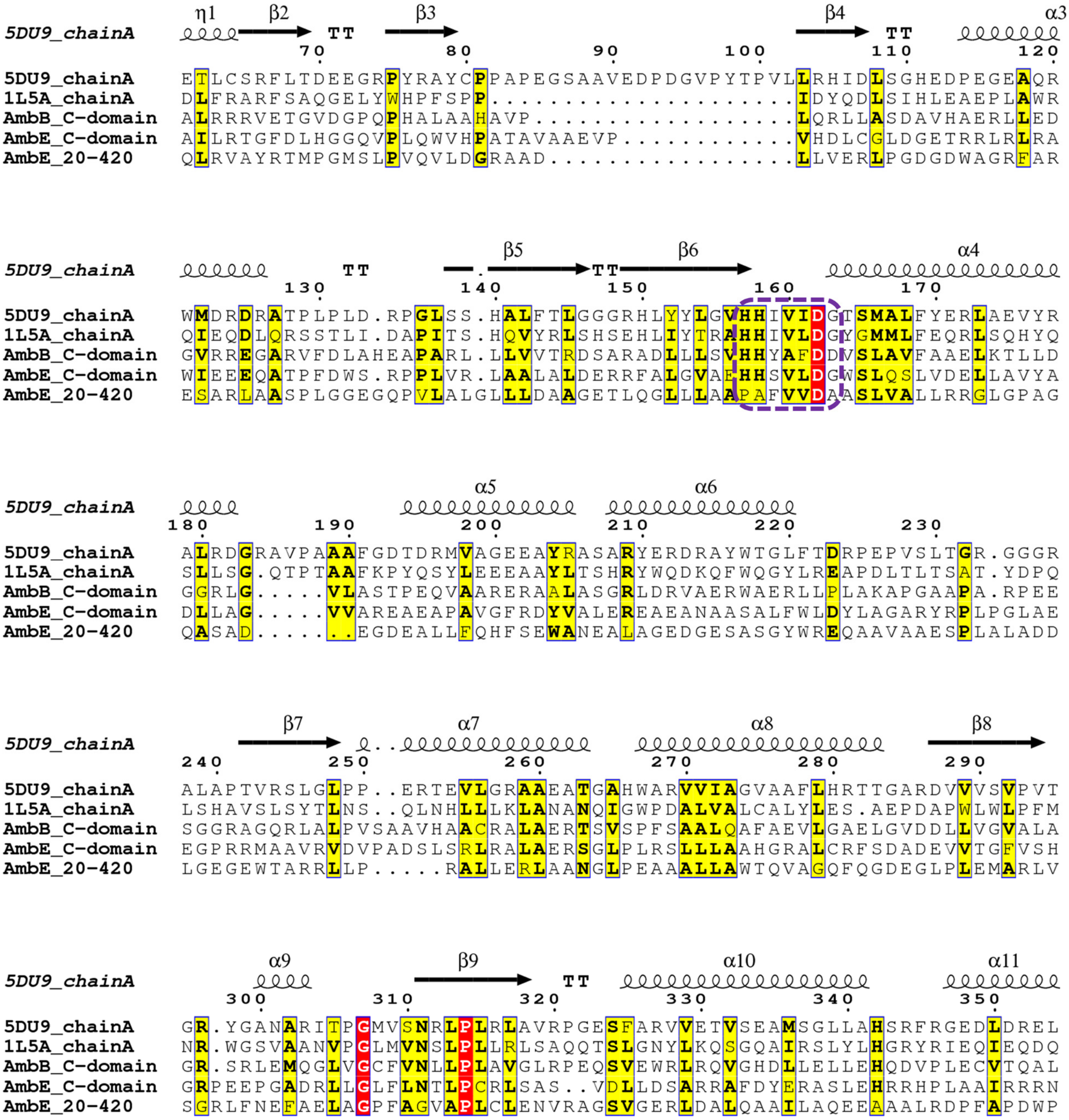


Fig 4-71: Protein sequence alignment of different condensation domains and the uncharacterized domain of AmbE. 5DU9 and 1L5A are deposited structures in the PDB (Bloudoff *et al.*, 2016; Keating *et al.*, 2002). The sequences of condensation (C) domain of AmbB and AmbE were identified and selected based on the prediction result of InterPro online server (Hunter *et al.*, 2009). This figure was generated with ESPript (Robert & Gouet, 2014) and adjusted manually to optimize the structure-based alignment. Secondary structure elements and turns (TT) of 5DU9 are shown. Identical residues are boxed and colored in red, and the conserved motif sequence HHXXXDG that is absent in AmbE 20-420 is circled by a dashed purple frame, While the motif is changed slightly to HHXXXDD in the condensation domain of AmbB.

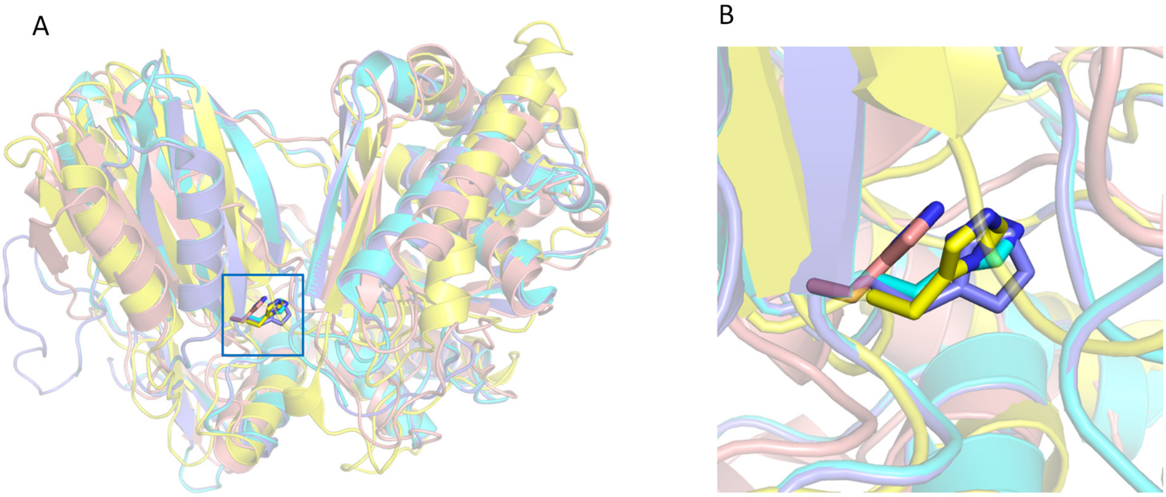


Fig 4-72: Superposition of four different condensation domains. A: Four condensation domain structures were 5DU9, 1L5A, 2JGP and 4JN3 colored in cyan, yellow, salmon and slate respectively (Bloudoff *et al.*, 2016; Bloudoff *et al.*, 2013; Keating *et al.*, 2002). B: The second Histidine in motif sequence HHXXXDG was zoomed in and shown as sticks (carbon atoms were colored accordingly with the cartoon).

In cyclization domains, the HHXXXDG sequence motif is replaced by DXXXXD, and the conserved Asp residues are critical for both condensation and heterocyclization (Keating *et al.*, 2000). An alignment between cyclization and condensation domains is shown in appendix figure 6-3. Again, no DXXXXD motif can be found in the sequence of AmbE 20-420. In addition, the final product AMB does not contain a circular group, such that the uncharacterized AmbE 20-420 domain cannot be a cyclization domain either.

In conclusion, the top hits obtained with DALI and PDBeFold do not lead to clear hypotheses about the function of AmbE 20-420. In another study, it has been reported that a so-called "X-domain", which was structurally also related to NRPS condensation domains, recruited free tailoring enzymes (Haslinger *et al.*, 2015). Mixture proteins of AmbE, AmbE 20-420, AmbC and AmbD were prepared and the size-exclusion chromatography (SEC) experiment were performed to preliminary check whether the uncharacterized domain AmbE 20-420 could recruit the tailoring enzymes AmbC and AmbD, but no indications for the complex formation have been found (appendix Fig 6-4). Therefore, the function of the uncharacterized domain AmbE 20-420 needs more exploration.

5. Conclusion and outlook

5.1 Fe/α-ketoglutarate-dependent dioxygenases AmbC and AmbD

In this study, Fe/α-ketoglutarate-dependent dioxygenase AmbC and AmbD from *Pseudomonas aeruginosa* could be expressed heterologously in *E. coli* and purified in large amounts with high purity. The crystal structures of apo AmbC and in complex with co-substrate analog NOG were determined, respectively. The flexible loop located above the active site was absent in the structure of apo AmbC, while otherwise no significant differences were noted between the two structures. AmbC adopts a closed state in the complex with NOG, with a channel extending to the metal binding active site. The depth of this channel was more or less equal to the length of a phosphopantetheine arm, suggesting that oxygenation at the *L*-Glu moiety occurs in an intermediate that is tethered to the thiolation (T) domain by a phosphopantetheine arm at the end of the channel.

The structure of AmbD in complex with co-substrate α-ketoglutarate also represents a closed state with a channel similar to that of AmbC/NOG. The structure of apo AmbD showed an opened state and did not contain a divalent metal cation in its active site. Instead, the conserved and metal-coordinating residue Glu^{124} interacted with Arg^{100}, which normally binds the α-ketoglutarate co-substrate to somehow pull the loop towards the active site.

Furthermore, by comparing the two complex structures, it was found that a bigger tunnel near the metal binding active site is present in AmbD than in AmbC. This may indicate that oxygenation on *L*-Glu moiety performed by AmbD occurs after that by AmbC. Several potential substrate binding residues could be identified by comparing the complex structures with available Fe/α-ketoglutarate-dependent dioxygenases (CarC and BBOX1) which contain the specific substrates.

5.2 Uncharacterized domain of AmbE

The optimized fragment AmbE 20-420 was designed to explore the function of this uncharacterized domain. Native and SeMet-labelled proteins were purified and crystallization screens were performed. The final structure of AmbE 20-420 was determined at a resolution of 2.1 Å. The overall architecture analysis by DALI and PDBeFold servers suggested a high similarity to NRPS condensation or cyclization domains. However, the conserved sequence motifs HHXXXDG or DXXXXD were absent in AmbE 20-420 and further, the two other condensation domains within AmbB and AmbE are sufficient to synthesize a tripeptide product, such that a third condensation domain seems unnecessary. As for cyclization, there is no circular moiety in the final product AMB, such that it is unlikely that the domain has this function either.

5.3 Outlook

In the course of this thesis work, the structures of dioxygenases AmbC and AmbD as apo enzymes and also in complex with the co-substrate or an analog have been obtained. The specific substrates are still unknown. From the initial substrates (two *L*-Ala and one *L*-Glu) to the final AMB molecule, many reactions have to occur. It can be hypothesized that AmbC and AmbD, as dioxygenases, will hydroxylate the β- and γ-positions of the *L*-Glu moiety selectively. Then, a methyl group from S-adenosyl methionine (SAM) will be transferred to the γ-hydroxyl group by the methyltransferase domain of AmbE. Dehydrogenation and decarboxylation would follow to synthesize the final AMB moiety, but it remains unclear which domains or parts might be responsible for these two conversions. At present, it is also not possible to decide whether these reactions or some of them occur to *L*-Glu, *L*-Ala-*L*-Glu dipeptide or *L*-Ala-*L*-Glu-*L*-Ala tripeptide. A hypothetical synthesis model proposed by our cooperator (Prof. Chris Calderone, Carleton College, USA) is shown in Fig 5-1. Just for clarity, it was assumed for this model that all modifications happen to the *L*-Ala-*L*-Glu dipeptide.

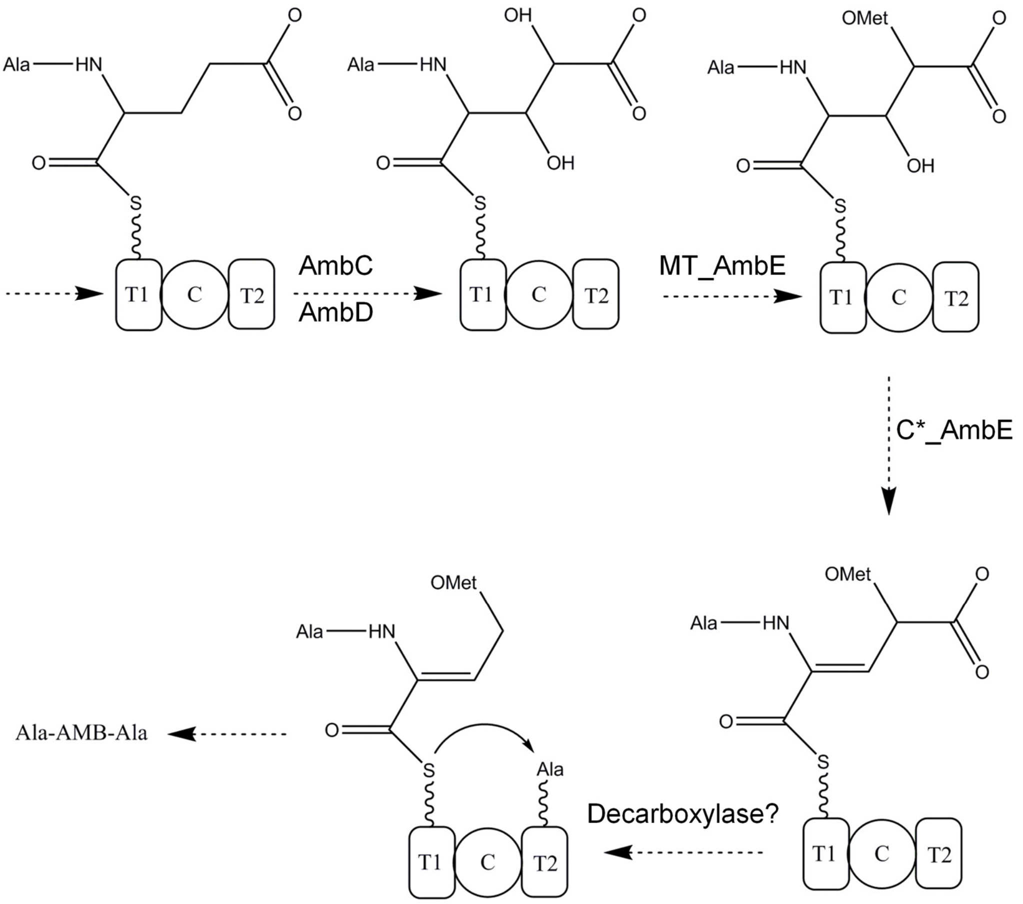


Fig 5-1: Modified model for AMB synthesis originally proposed by Chris Calderone (data are not published). For clarity, all modifications of the *L*-Glu moiety are assumed to occur to the dipepetide intermediate. AmbC and AmbD perform hydroxylation at the β- and γ-positions of *L*-Glu. The methyltransferase domain of AmbE adds a methyl group to the γ-hydroxyl group. Then by the C* domain or/and uncharacterized domain of AmbE, the tripeptide Ala-AMB-Ala is synthesized and released by the TE domain of AmbE.

For the uncharacterized domain of AmbE, it may play the role of a decarboxylase, an activity that is missing from the characterized domains of AmbB and AmbE. A truncated AmbE without this domain could be constructed to explore its function further.

Mass spectrometry method has been applied to explore how these proteins and functional domains work on different intermediates and the timing points of the tailoring reactions. By now, the activation and loading of the initial substrate *L*-Glu moiety to the T1 domain of mutated AmbE (T2 domain mutant) could be observed. For the next step, by controlling the presence of AmbC, AmbD, Fe^{2+}/α-ketoglutarate and SAM, useful clues about the intermediates and timing points of the reactions can be deduced. With the similar way, the dipeptide *L*-Ala-*L*-Glu can be formed under the presence of AmbB and AmbE (T2 domain mutant). Also, the tripeptide *L*-Ala-*L*-Glu-*L*-Ala can be generated under the function of AmbB and AmbE (TE domain mutant). With all the results, the intermediates of different enzymes or functional domains could be deduced, as well as the timing points of reactions.

Then, the substrate binding can be clarified by determining the complex structures of AmbC and that of AmbD. Furthermore, the reaction mechanism will be unveiled.

6. Appendix

6.1 Information about 20 amino acids

Table 6-1: The proteinogenic amino acids

One letter	Three letter	Amino acid	One letter	Three letter	Amino acid
A	Ala	alanine	M	Met	methionine
C	Cys	cysteine	N	Asn	asparagine
D	Asp	aspartate	P	Pro	proline
E	Glu	glutamate	Q	Gln	glutamine
F	Phe	phenylalanine	R	Arg	arginine
G	Gly	glycine	S	Ser	serine
H	His	histidine	T	Thr	threonine
I	Ile	isoleucine	V	Val	valine
K	Lys	lysine	W	Trp	tryptophan
L	Leu	leucine	Y	Tyr	tyrosine

6.2 DNA-sequences and related translations in this study

Table 6-2: DNA and protein sequences of AmbC

AmbC 30-350 is highlighted in grey

```
  1 ATGGAACGAACAGCTCCCAGTCTCACCGGTCGCGGCTTCGCCCGGCCGCGCAATCTCAGC 60
  1  M  E  R  T  A  P  S  L  T  G  R  G  F  A  R  P  R  N  L  S  20

 61 CCGCGCAGCAGCGAGGCGCTGGTGCGGCGCCCGGCGGCGGAGCGCGACGGCCTGCCGCTG 120
 21  P  R  S  S  E  A  L  V  R  R  P  A  A  E  R  D  G  L  P  L  40

121 CTGATCGAGGCACAGGCTCCCGGCCTGTCCCTGGCCGACTGGATTCGCGAGCAGGGCCAG 180
 41  L  I  E  A  Q  A  P  G  L  S  L  A  D  W  I  R  E  Q  G  Q  60

181 TCCCTGCACGACGACCTGAACCTGGCTGGCGGCCTGTTGCTGAGAGGCTTCGAGGTCGAC 240
 61  S  L  H  D  D  L  N  L  A  G  G  L  L  L  R  G  F  E  V  D  80

241 TCGGCGGAGCGCTTCCGCGCAGCCGCCGCTGCCTTCGCCCCGCAACTGCTCGACTACAAG 300
```

```
  81  S   A   E   R   F   R   A   A   A   A   A   F   A   P   Q   L   L   D   Y   K  100

 301 GAGCGTTCCTCGCCGCGTAGCCAGGTCAGCGGCGAGGTCTACACCTCCACCGAGCATCCG 360
 101  E   R   S   S   P   R   S   Q   V   S   G   E   V   Y   T   S   T   E   H   P  120

 361 GTGGACCAGCCGATCTTCCTGCATAACGAGCAGTCCTATACCGCCGACTGGCCGCTGTAC 420
 121  V   D   Q   P   I   F   L   H   N   E   Q   S   Y   T   A   D   W   P   L   Y  140

 421 ATCATGTTCCATTGCCAGGTCGCGCCGCTCGAGGGCGGCGCCACTCCGGTGGCGGCCAAC 480
 141  I   M   F   H   C   Q   V   A   P   L   E   G   G   A   T   P   V   A   A   N  160

 481 CGGCTGGTGCTCCGCCACCTGCCGGACGAATTGCTGGAACGCTTCGGGCGCCTGGGCATC 540
 161  R   L   V   L   R   H   L   P   D   E   L   L   E   R   F   G   R   L   G   I  180

 541 CTCTACGTGCGCAACTACCGCGCCGGCCTCGGCCTGTCCTGGCGCGAAGCGTTCCAGACC 600
 181  L   Y   V   R   N   Y   R   A   G   L   G   L   S   W   R   E   A   F   Q   T  200

 601 GACAGCCGCGCCGAGGTCGAGGCGTTCTGCGCTGAGCACCGGATCGCCCATGCCTGGATC 660
 201  D   S   R   A   E   V   E   A   F   C   A   E   H   R   I   A   H   A   W   I  220

 661 GGCGACGAGCACCTGCGCACCTGGCAGAGGCGCGCGGCGTTCCAACGCCACCCGTATACC 720
 221  G   D   E   H   L   R   T   W   Q   R   R   A   A   F   Q   R   H   P   Y   T  240

 721 GGCGAGCGCCTGTGGTTCAACCACGGCATGTTCTTCCATGCCACCAGTCTCGAGCCGGGC 780
 241  G   E   R   L   W   F   N   H   G   M   F   F   H   A   T   S   L   E   P   G  260

 781 CTGCGCGATGCGCTGCTGCGCAGCGTCGCCGAGGAAGACCTGCCGTACCAGACCTACTAC 840
 261  L   R   D   A   L   L   R   S   V   A   E   E   D   L   P   Y   Q   T   Y   Y  280

 841 GGCGATGGCAGCCCGATCGAGGCGCAGACCCTGGCGACCATCCGCAGCGCCATAGACCGC 900
 281  G   D   G   S   P   I   E   A   Q   T   L   A   T   I   R   S   A   I   D   R  300

 901 GAGACGCGGCGCTTCGACTGGCGCGTGGGCGACGTGCTGATCCTCGACAACATGCTCGCC 960
 301  E   T   R   R   F   D   W   R   V   G   D   V   L   I   L   D   N   M   L   A  320

 961 CAGCACGGCCGCGAGCCGTTCCGCGGGCCGCGACGGATCCTCACCATCATGTCCACGCCG 1020
 321  Q   H   G   R   E   P   F   R   G   P   R   R   I   L   T   I   M   S   T   P  340

1021 TATTCCAGCCTGGCGCCCGCCGACGTCCCGCCGACACCTCCGCGCCTGGCGCTCGGAGGT 1080
 341  Y   S   S   L   A   P   A   D   V   P   P   T   P   P   R   L   A   L   G   G  360

1081 GCCGCATGA 1089
 361  A   A   *   363
```

Table 6-3: DNA and protein sequences of AmbD

```
   1 ATGAGCGCCTCGTTCAGCGCGCCGCGCCTGCGCCCCCGGCAGCTCTCGGCCGGCGACCTG 60
   1  M   S   A   S   F   S   A   P   R   L   R   P   R   Q   L   S   A   G   D   L  20

  61 GTGGAGGAAAGCCTGCTCGACCCGGCCAACGACTATCTGCGCATCGTCCGCGCGCGCCAG 120
```

```
 21  V  E  E  S  L  L  D  P  A  N  D  Y  L  R  I  V  R  A  R  Q  40

121 CCGGGCATGGACCTGCGGCAATGGATCGCCGCCGCCGGCGCAGGGCTGCGCGACAGCCTG 180
 41  P  G  M  D  L  R  Q  W  I  A  A  A  G  A  G  L  R  D  S  L  60

181 CTGCGGCATGGCGGCATCCTGTTTCGTGGCTTCGCGGTAGACGGAGCGGAGGGCTTCTCC 240
 61  L  R  H  G  G  I  L  F  R  G  F  A  V  D  G  A  E  G  F  S  80

241 CAGGCGGTACAAAGCTTTTCGCCGAACATGCTCGACTACCTGGAGCGCGCGGCGGCACGC 300
 81  Q  A  V  Q  S  F  S  P  N  M  L  D  Y  L  E  R  A  A  A  R  100

301 CAGGAAGTGGCGCACCGGGTGTTCACCTCCACCGAGTTCTCGCCGGACGGCTGGATTCCG 360
101  Q  E  V  A  H  R  V  F  T  S  T  E  F  S  P  D  G  W  I  P  120

361 CCGCACCACGAGATGTCCTACTCGCACAACTGGCCGAGCTATATCCACTTCTACTGCCAG 420
121  P  H  H  E  M  S  Y  S  H  N  W  P  S  Y  I  H  F  Y  C  Q  140

421 ACACCGCCCGCCACCCAGGGCCGCACGCCCCTGGCCGACGAGCGCCGGGTCAGCGCTCGG 480
141  T  P  P  A  T  Q  G  R  T  P  L  A  D  E  R  R  V  S  A  R  160

481 ATACCCGAGGCGATCAGGCAGCGCTTCCTCCGCCATGGGGTGTGCTACGTGCGCAACTAC 540
161  I  P  E  A  I  R  Q  R  F  L  R  H  G  V  C  Y  V  R  N  Y  180

541 GGGCCGGAGATCGACCTGACCTGGCAGGAAGGCTTCCAGACCGACAGTCGCGCCGAGGTC 600
181  G  P  E  I  D  L  T  W  Q  E  G  F  Q  T  D  S  R  A  E  V  200

601 GAGGCCTACTGCCGGCAGACCGGTACCCAATGGACCTGGCTCGACGACCAGCGCCTGAAT 660
201  E  A  Y  C  R  Q  T  G  T  Q  W  T  W  L  D  D  Q  R  L  N  220

661 ACCCGCCAGGTACGCCAGGCGATGGTGCGCCACCCGCTCAGCGGCGAGACGCTGTGGTTC 720
221  T  R  Q  V  R  Q  A  M  V  R  H  P  L  S  G  E  T  L  W  F  240

721 AACCACGCGCACATGTTCCATGTCTCCAACATGCCGCCGGCCCTGGCCCGGGCCTTGCTC 780
241  N  H  A  H  M  F  H  V  S  N  M  P  P  A  L  A  R  A  L  L  260

781 GACGAGGTCGGCGAACAGGGCCTGCCGCGCAACGCCTACTACGGCGACGGCAGCCCGATC 840
261  D  E  V  G  E  Q  G  L  P  R  N  A  Y  Y  G  D  G  S  P  I  280

841 GAGGCGGAGGTGCTGGACACCATCCGCGCCGCCTACCGCGAGGAAACCCGCGCGTTCGCC 900
281  E  A  E  V  L  D  T  I  R  A  A  Y  R  E  E  T  R  A  F  A  300

901 TGGGAGCGCGGCGATGTCCTGATGCTCGACAACTTCATCAGCGTCCACGGCCGCGAGCCG 960
301  W  E  R  G  D  V  L  M  L  D  N  F  I  S  V  H  G  R  E  P  320

961 TACACCGGCGAGCGCAAGGTGCTGGTGGCGATGACCGACCTGCATGTCCACCAACCCTGA 1020
321  Y  T  G  E  R  K  V  L  V  A  M  T  D  L  H  V  H  Q  P  *  340
```

Table 6-4: DNA and protein sequences of AmbE 20-420

```
  1 GAAGGATTCCCGCTGTCTCCCTTGCAGACCCGCGCCTGGCGCCGCCATGCCGAGCGGCCG 60
  1  E  G  F  P  L  S  P  L  Q  T  R  A  W  R  R  H  A  E  R  P  20
```

```
  61 GAAAATACGGTTGTCGGCGTGCGCCTGCACGCCCCGGCCGATCCCGTGGCGACGCTGGAG 120
  21  E  N  T  V  V  G  V  R  L  H  A  P  A  D  P  V  A  T  L  E  40

 121 CGGCTGCGCCGGGCGCTGGACGGCGAGGCGCAACTGCGCGTGGCCTACCGGACGATGCCG 180
  41  R  L  R  R  A  L  D  G  E  A  Q  L  R  V  A  Y  R  T  M  P  60

 181 GGCATGAGCCTGCCGGTGCAGGTACTGGATGGGCGCGCGGCCGATCTGCTGGTCGAGCGC 240
  61  G  M  S  L  P  V  Q  V  L  D  G  R  A  A  D  L  L  V  E  R  80

 241 CTGCCGGGAGACGGCGACTGGGCCGGACGCTTCGCGCGCGAAAGCGCGCGTCTCGCCGCT 300
  81  L  P  G  D  G  D  W  A  G  R  F  A  R  E  S  A  R  L  A  A  100

 301 TCGCCCCTGGGCGGGGAAGGCCAGCCGGTACTGGCGCTCGGCCTGCTGCTGGACGCCGCC 360
 101  S  P  L  G  G  E  G  Q  P  V  L  A  L  G  L  L  L  D  A  A  120

 361 GGAGAGACGCTCCAGGGGCTGTTGCTGGCGGCGCCGGCGTTCGTCGTCGATGCGGCCAGC 420
 121  G  E  T  L  Q  G  L  L  L  A  A  P  A  F  V  V  D  A  A  S  140

 421 CTGGTGGCGCTGCTGCGCCGCGGCCTGGGGCCGGCCGGCCAGGCGAGCGCGGACGAGGGA 480
 141  L  V  A  L  L  R  R  G  L  G  P  A  G  Q  A  S  A  D  E  G  160

 481 GACGAGGCGCTGCTGTTCCAGCATTTCTCCGAGTGGGCCAACGAGGCGCTGGCCGGCGAA 540
 161  D  E  A  L  L  F  Q  H  F  S  E  W  A  N  E  A  L  A  G  E  180

 541 GACGGCGAAAGCGCCAGCGGTTACTGGCGAGAGCAGGCGGCCGTTGCGGCGGAGAGTCCG 600
 181  D  G  E  S  A  S  G  Y  W  R  E  Q  A  A  V  A  A  E  S  P  200

 601 CTGGCGCTGGCGGACGACCTGGGCGAAGGCGAGTGGACGGCGCGGCGCCTGCTGCCGCGC 660
 201  L  A  L  A  D  D  L  G  E  G  E  W  T  A  R  R  L  L  P  R  220

 661 GCGCTGCTCGAACGCCTGGCCGCCAACGGCTTGCCGGAGGCGGCCGCCCTGCTGGCCTGG 720
 221  A  L  L  E  R  L  A  A  N  G  L  P  E  A  A  A  L  L  A  W  240

 721 ACCCAGGTCGCCGGGCAGTTCCAGGGCGACGAGGGCCTCCCGCTGGAAATGGCGCGACTG 780
 241  T  Q  V  A  G  Q  F  Q  G  D  E  G  L  P  L  E  M  A  R  L  260

 781 GTCTCGGGGCGCCTGTTCAACGAGTTCGCCGAGCTGGCCGGACCGTTCGCCGGGGTCGCG 840
 261  V  S  G  R  L  F  N  E  F  A  E  L  A  G  P  F  A  G  V  A  280

 841 CCGCTGTGCCTGGAGAATGTCCGCGCGGGCAGCGTCGGCGAGCGGCTCGACGCCCTCCAG 900
 281  P  L  C  L  E  N  V  R  A  G  S  V  G  E  R  L  D  A  L  Q  300

 901 GCGGCGATCCTCGCCCAGGAGGAGGCAGCGGCCCTGCGCGATCCCTTTGCCCCCGACTGG 960
 301  A  A  I  L  A  Q  E  E  A  A  A  L  R  D  P  F  A  P  D  W  320

 961 CCGCTCGCCGAGTTGGGCTTCGCCTGGCTGGCGGGCGAACTGGATGGCGCCGGGGTGGCC 1020
 321  P  L  A  E  L  G  F  A  W  L  A  G  E  L  D  G  A  G  V  A  340

1021 GAGCTGGATTGCCGTCAGCCGCCGCTGGGCGGGTTCCTCGAGTTGCAGGTGCTGCCCCAC 1080
 341  E  L  D  C  R  Q  P  P  L  G  G  F  L  E  L  Q  V  L  P  H  360

1081 GGCGAAGGCAGGCTGGCCAGCCTGCGGGTCCGTCGCGACCATGACGGAACGCTGGCCGGG 1140
 361  G  E  G  R  L  A  S  L  R  V  R  R  D  H  D  G  T  L  A  G  380
```

```
1141 CGCTTGCTCGACGCCTGGGTCGAATGCCTGGAAAGCATCGCCGCCGACAGGCAACTGCCA 1200
 381  R  L  L  D  A  W  V  E  C  L  E  S  I  A  A  D  R  Q  L  P  400

1201 CTG 1203
 401  L  401
```

6.3 Primers used in this study

Table 6-5 Primers used in this study

Name	Sequence 5' to 3'
ambC_F	CTGGAAGTTCTGTTTCAGGGTACCATGGAACGAACAGCTCCCAGTCTC
ambC_R	GGTGTTTAAATGGTCTAGAAAGCTTATGCGGCACCTCCGAGC
ambD_F	CTGGAAGTTCTGTTTCAGGGTACCATGAGCGCCTCGTTCAGC
ambD_R	GGTGTTTAAATGGTCTAGAAAGCTTAGGGTTGGTGGACATGCAGG
AmbC 30_F	CTGGAAGTTCTGTTTCAGGGTACCCGCCCGGCGGCGGAGCGCGA
ambC 350_R	GGTGTTTAAATGGTCTAGAAAGCTTACGGGACGTCGGCGGGCGCCA
ambD 30_F	CTGGAAGTTCTGTTTCAGGGTACC AACGACTATCTGCGCATCGTCC
ambD 334_R	GGTGTTTAAATGGTCTAGAAAGCTTACAGGTCGGTCATCGCCACCA
AmbE 20_F	CTGGAAGTTCTGTTTCAGGGTACCGAAGGATTCCCGCTGTCTCCCTT
ambE 440_R	GGTGTTTAAATGGTCTAGAAAGCTTACTCGCCCTGCCAGGCCT
ambE 420_R	GGTGTTTAAATGGTCTAGAAAGCTTAAGTGGCAGTTGCCTGTCGGC

6.4 Additional figures

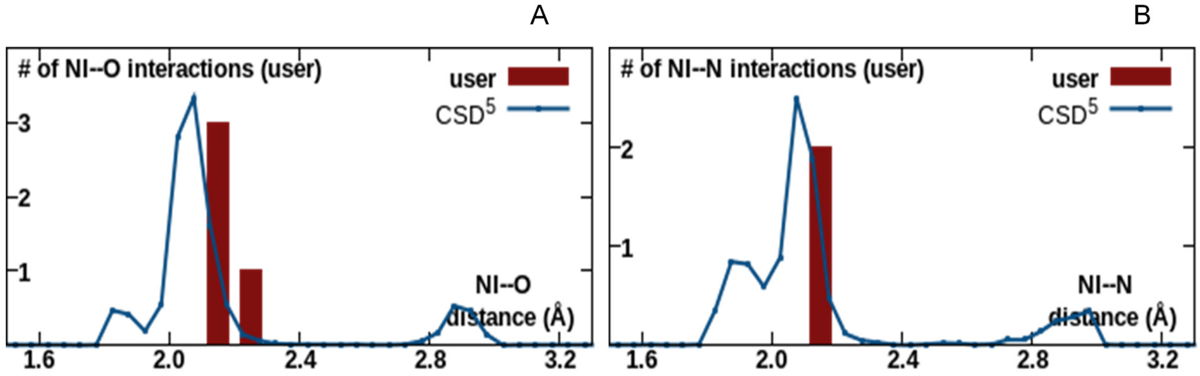


Fig 6-1: Nickel metal-ligand distance distribution in the AmbC 30-350 with NOG structure compared with CSD. A: Ni-oxygen (four in total) distances. B: Ni-nitrogen (two in total) distances. The blue curve represents the distance distribution from CSD, a database of very high-resolution X-ray crystallography structures of small molecules. User means the submitted structure.

Table 6-6: Chosen parameters generated from CMM report of complex structure of AmbC 30-350 with NOG

ID	Metal	B factor (env.)	Ligands	valence	Geometry	gRMSD (°)	Vancancy
D: 1	NI	12.6 (11.7)	O_4N_2	1.7	Octahedral	5.9°	0

Ligand: Not applicable Outlier *Borderline* Acceptable

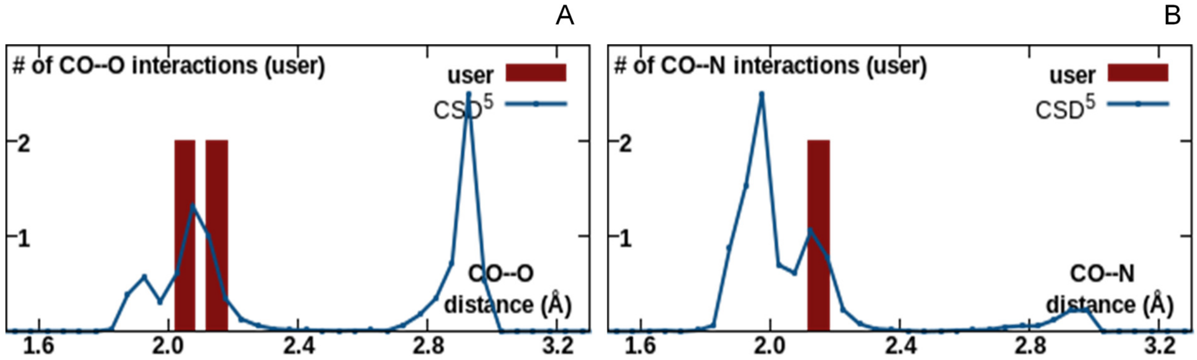
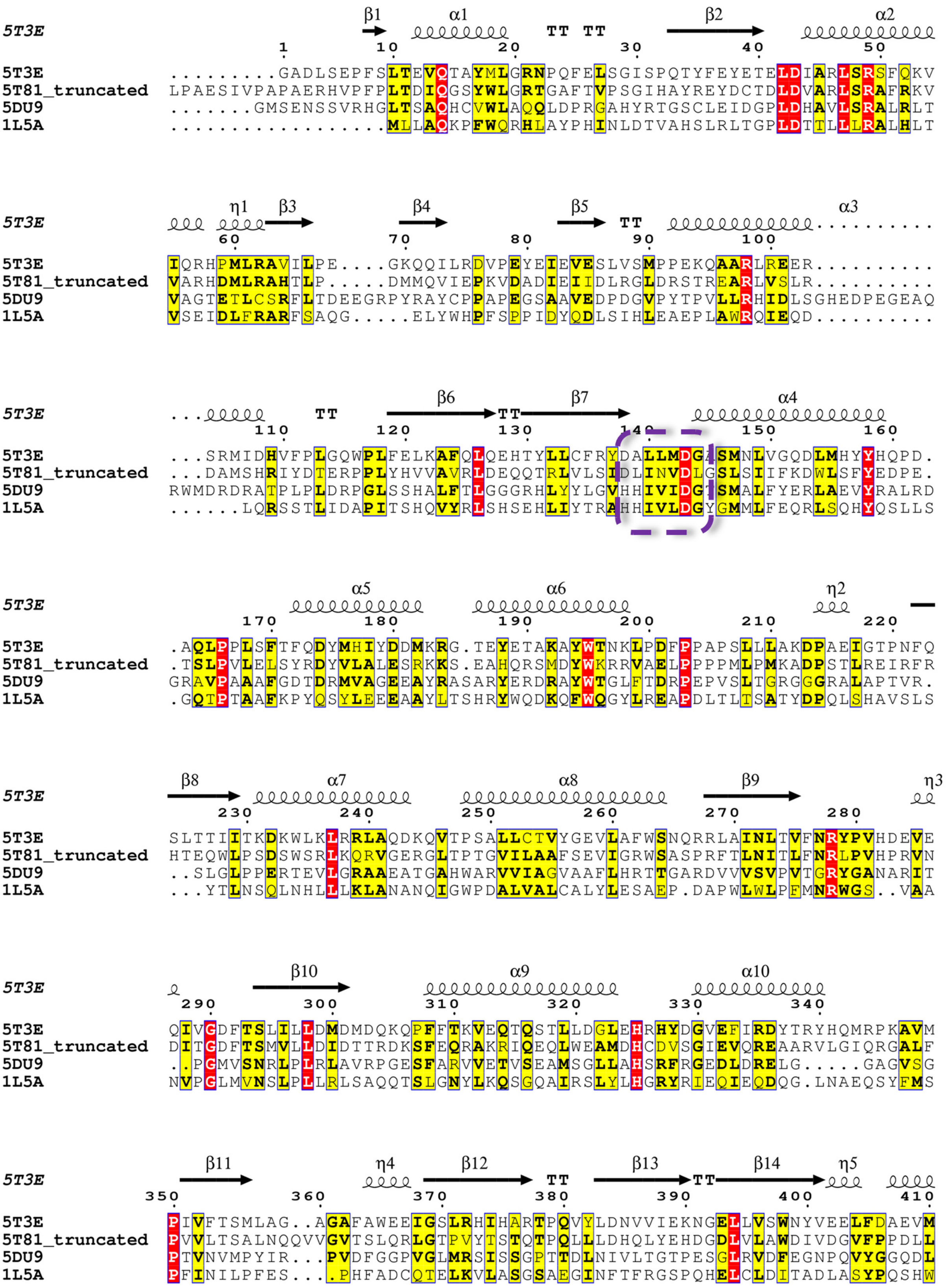


Fig 6-2: Cobalt metal-ligand distance distribution in the structure of AmbD in complex with α-KG compared with CSD. A: Co-oxygen (four in total) distances. B: Co-nitrogen (two in total) distances. The blue curve represents the distance distribution from CSD, a database of very high-resolution X-ray crystallography structures of small molecules. User means the submitted structure.

Table 6-7: Chosen parameters generated from CMM report of complex structure of AmbD with α-KG.

ID	Metal	B factor (env.)	Ligands	valence	Geometry	gRMSD (°)	Vancancy
D: 1	CO	14.7 (15.7)	O_4N_2	1.9	Octahedral	5.4°	0

Ligand: Not applicable Outlier *Borderline* Acceptable



```
5T3E            .........GADLSEPFSLTEVQTAYMLGRNPQFELSGISPQTYFEYETELDIARLSRSFQKV
5T81_truncated  LPAESIVPAPAERHVPFPLTDIQGSYWLGRTGAFTVPSGIHAYREYDCTDLDVARLSRAFRKV
5DU9            .......GMSENSSVRHGLTSAQHCVWLAQQLDPRGAHYRTGSCLEIDGPLDHAVLSRALRLT
1L5A            ..................MLLAQKPFWQRHLAYPHINLDTVAHSLRLTGPLDTTLLLRALHLT

5T3E            IQRHPMLRAVILPE....GKQQILRDVPEYEIEVESLVSMPPEKQAARLREER..........
5T81_truncated  VARHDMLRAHTLP.....DMMQVIEPKVDADIEIIDLRGLDRSTREARLVSLR..........
5DU9            VAGTETLCSRFLTDEEGRPYRAYCPPAPEGSAAVEDPDGVPYTPVLLRHIDLSGHEDPEGEAQ
1L5A            VSEIDLFRARFSAQG....ELYWHPFSPPIDYQDLSIHLEAEPLAWRQIEQD..........

5T3E            ...SRMIDHVFPLGQWPLFELKAFQLQEHTYLLCFRYDALLMDGASMNLVGQDLMHYYHQPD.
5T81_truncated  ...DAMSHRIYDTERPPLYHVVAVRLDEQQTRLVLSIDLINVDLGSLSIIFKDWLSFYEDPE.
5DU9            RWMDRDRATPLPLDRPGLSSHALFTLGGGRHLYYLGVHHIVIDGTSMALFYERLAEVYRALRD
1L5A            ......LQRSSTLIDAPITSHQVYRLSHSEHLIYTRAHHIVLDGYGMMLFEQRLSQHYQSLLS

5T3E            .AQLPPLSFTFQDYMHIYDDMKRG.TEYETAKAYWTNKLPDFPPAPSLLLAKDPAEIGTPNFQ
5T81_truncated  .TSLPVLELSYRDYVLALESRKKS.EAHQRSMDYWKRRVAELPPPPMLPMKADPSTLREIRFR
5DU9            GRAVPAAAFGDTDRMVAGEEAYRASARYERDRAYWTGLFTDRPEPVSLTGRGGGRALAPTVR.
1L5A            .GQTPTAAFKPYQSYLEEEAAYLTSHRYWQDKQFWQGYLREAPDLTLTSATYDPQLSHAVSLS

5T3E            SLTTIITKDKWLKLRRLAQDKQVTPSALLCTVYGEVLAFWSNQRRLAINLTVFNRYPVHDEVE
5T81_truncated  HTEQWLPSDSWSRLKQRVGERGLTPTGVILAAFSEVIGRWSASPRFTLNITLFNRLPVHPRVN
5DU9            ..SLGLPPERTEVLGRAAEATGAHWARVVIAGVAAFLHRTTGARDVVVSVPVTGRYGANARIT
1L5A            ...YTLNSQLNHLLLKLANANQIGWPDALVALCALYLESAEP.DAPWLWLPFMNRWGS..VAA

5T3E            QIVGDFTSLILLDMDMDQKQPFFTKVEQTQSTLLDGLEHRHYDGVEFIRDYTRYHQMRPKAVM
5T81_truncated  DITGDFTSMVLLDIDTTRDKSFEQRAKRIQEQLWEAMDHCDVSGIEVQREAARVLGIQRGALF
5DU9            ..PGMVSNRLPLRLAVRPGESFARVVETVSEAMSGLLAHSRFRGEDLDRELG.....GAGVSG
1L5A            NVPGLMVNSLPLLRLSAQQTSLGNYLKQSGQAIRSLYLHGRYRIEQIEQDQG.LNAEQSYFMS

5T3E            PIVFTSMLAG..AGAFAWEEIGSLRHIHARTPQVYLDNVVIEKNGELLVSWNYVEELFDAEVM
5T81_truncated  PVVLTSALNQQVVGVTSLQRLGTPVYTSTQTPQLLLDHQLYEHDGDLVLAWDIVDGVFPPDLL
5DU9            PTVNVMPYIR...PVDFGGPVGLMRSISSGPTTDLNIVLTGTPESGLRVDFEGNPQVYGGQDL
1L5A            PFINILPFES....PHFADCQTELKVLASGSAEGINFTFRGSPQHELCLDITADLASYPQSHW
```

Fig 6-3: Protein sequence alignment of condensation and cyclization domains. 5DU9, 1L5A, 5T3E and 5T81 are deposited structures on PDB. The former two structures are condensation domains and the latter two are cyclization domains. This figure was generated with ESPript (Robert & Gouet, 2014) and adjusted manually to optimize structure-based alignment. Secondary structure elements and turns (TT) of 5T3E were shown. Strictly conserved residues were boxed and colored in red, and the conserved motif sequence HHXXXDG in condensation domains was replaced by DXXXXD in cyclization domains which was circled by a dashed purple frame. For clarity, some residues from C terminals were not shown.

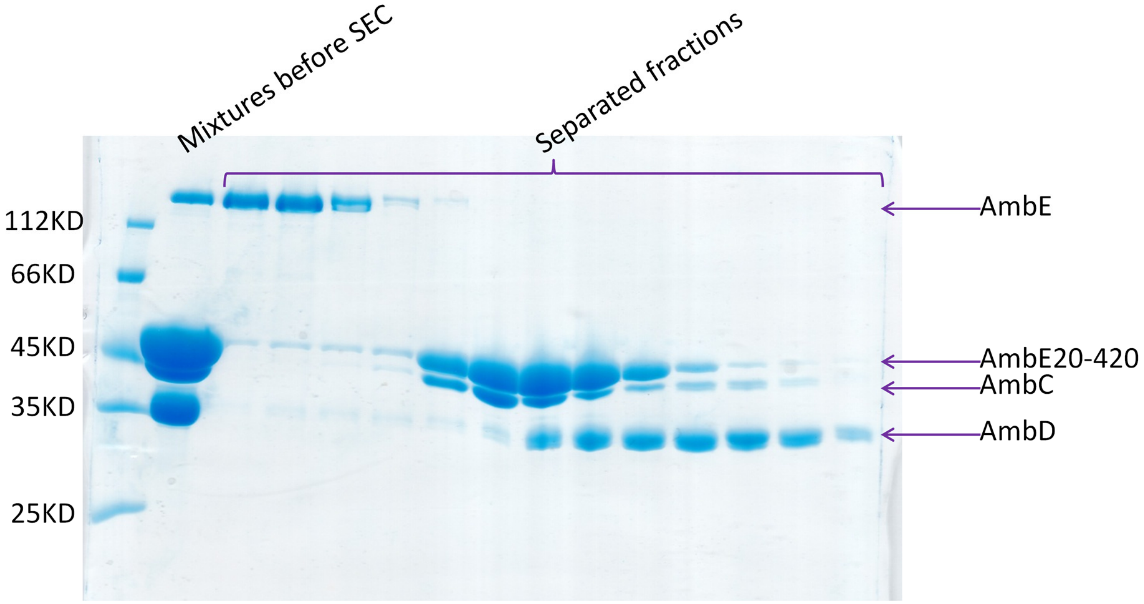


Fig 6-4: SDS-PAGE followed by InstantBlue staining for size-exclusion chromatography of AmbE, AmbE 20-420, AmbC and AmbD protein mixture. Based on molecular weights of different proteins, no protein complex could be identified from all the fractions.

7. References

Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., . . . Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 2), 213-221.

Alexander McPherson, J. W. S. (1982). *Preparation and analysis of protein crystals*.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol, 215*(3), 403-410.

Arancibia, F., Bauer, T. T., Ewig, S., Mensa, J., Gonzalez, J., Niederman, M. S., & Torres, A. (2002). Community-acquired pneumonia due to gram-negative bacteria and pseudomonas aeruginosa: incidence, risk, and prognosis. *Arch Intern Med, 162*(16), 1849-1858.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A, 98*(18), 10037-10041.

Balibar, C. J., Vaillancourt, F. H., & Walsh, C. T. (2005). Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem Biol, 12*(11), 1189-1200.

Bergendahl, V., Linne, U., & Marahiel, M. A. (2002). Mutational analysis of the C-domain in nonribosomal peptide synthesis. *Eur J Biochem, 269*(2), 620-629.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., . . . Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res, 28*(1), 235-242.

Biggins, J. B., Ternei, M. A., & Brady, S. F. (2012). Malleilactone, a polyketide synthase-derived virulence factor encoded by the cryptic secondary metabolome of Burkholderia pseudomallei group pathogens. *J Am Chem Soc, 134*(32), 13192-13195.

Bitto, E., Bingman, C. A., Allard, S. T., Wesenberg, G. E., Aceti, D. J., Wrobel, R. L., . . . Phillips, G. N., Jr. (2005). The structure at 2.4 A resolution of the protein from gene locus At3g21360, a putative Fe(II)/2-oxoglutarate-dependent enzyme from Arabidopsis thaliana. *Acta Crystallogr Sect F Struct Biol Cryst Commun, 61*(Pt 5), 469-472.

Bloudoff, K., Alonzo, D. A., & Schmeing, T. M. (2016). Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases. *Cell Chem Biol, 23*(3), 331-339.

Bloudoff, K., Rodionov, D., & Schmeing, T. M. (2013). Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases. *J Mol Biol, 425*(17), 3137-3150.

Breidenstein, E. B., de la Fuente-Nunez, C., & Hancock, R. E. (2011). Pseudomonas aeruginosa: all roads lead to resistance. *Trends Microbiol, 19*(8), 419-426.

Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K., & Jones, D. T. (2013). Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res, 41*(Web Server issue), W349-357.

Bugg, T. D. H. (2003). Dioxygenase enzymes: catalytic mechanisms and chemical models. *Tetrahedron, 59*(36), 7075-7101.

Bunkoczi, G., Echols, N., McCoy, A. J., Oeffner, R. D., Adams, P. D., & Read, R. J. (2013). Phaser.MRage: automated molecular replacement. *Acta Crystallogr D Biol Crystallogr, 69*(Pt 11), 2276-2286.

Burkhardt, A. P., T. ; Reime, B. ; Meyer, J. ; Fischer, P. ; Stübe, N. ; Panneerselvam, S. ; Lorbeer, O. ; Stachnik, K. ; Warmer, M. ; Rödig, P. ; Göries, D. ; Meents, A. (2016). Status of the crystallography beamlines at PETRA III. *The European physical journal / Plus*(131), 1.

Bushley, K. E., Raja, R., Jaiswal, P., Cumbie, J. S., Nonogaki, M., Boyd, A. E., . . . Spatafora, J. W. (2013). The genome of tolypocladium inflatum: evolution, organization, and expression of the cyclosporin biosynthetic gene cluster. *PLoS Genet, 9*(6), e1003496.

Byford, M. F., Baldwin, J. E., Shiau, C. Y., & Schofield, C. J. (1997). The Mechanism of ACV Synthetase. *Chem Rev, 97*(7), 2631-2650.

Caboche, S., Leclere, V., Pupin, M., Kucherov, G., & Jacques, P. (2010). Diversity of monomers in nonribosomal peptides: towards the prediction of origin and biological activity. *J Bacteriol, 192*(19), 5143-5150.

Capitani, G., McCarthy, D. L., Gut, H., Grutter, M. G., & Kirsch, J. F. (2002). Apple 1-aminocyclopropane-1-carboxylate synthase in complex with the inhibitor L-aminoethoxyvinylglycine. Evidence for a ketimine intermediate. *J Biol Chem, 277*(51), 49735-49742.

Cardinale, G. J., Rhoads, R. E., & Udenfriend, S. (1971). Simultaneous incorporation of 18 O into succinate and hydroxyproline catalyzed by collagen proline hydroxylase. *Biochem Biophys Res Commun, 43*(3), 537-543.

Chang, W. C., Guo, Y., Wang, C., Butch, S. E., Rosenzweig, A. C., Boal, A. K., . . . Bollinger, J. M., Jr. (2014). Mechanism of the C5 stereoinversion reaction in the biosynthesis of carbapenem antibiotics. *Science, 343*(6175), 1140-1144.

Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., . . . Richardson, D. C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 1), 12-21.

Chernov, A. A. (2003). Protein crystals and their growth. *Journal of Structural Biology, 142*(1), 3-21.

Clifton, I. J., Doan, L. X., Sleeman, M. C., Topf, M., Suzuki, H., Wilmouth, R. C., & Schofield, C. J. (2003). Crystal structure of carbapenem synthase (CarC). *J Biol Chem, 278*(23), 20843-20850.

Clifton, I. J., Hsueh, L. C., Baldwin, J. E., Harlos, K., & Schofield, C. J. (2001). Structure of proline 3-hydroxylase. Evolution of the family of 2-oxoglutarate dependent oxygenases. *Eur J Biochem, 268*(24), 6625-6636.

Cornelis, P. (2008). *Pseudomonas: Genomics and Molecular Biology* (P. Cornelis Ed.): Caister Academic Press.

Cornell, N. W., Zuurendonk, P. F., Kerich, M. J., & Straight, C. B. (1984). Selective inhibition of alanine aminotransferase and aspartate aminotransferase in rat hepatocytes. *Biochem J, 220*(3), 707-716.

Correia, T., Grammel, N., Ortel, I., Keller, U., & Tudzynski, P. (2003). Molecular cloning and analysis of the ergopeptine assembly system in the ergot fungus Claviceps purpurea. *Chem Biol, 10*(12), 1281-1292.

Costas, M., Mehn, M. P., Jensen, M. P., & Que, L., Jr. (2004). Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates. *Chem Rev, 104*(2), 939-986.

Costas M, M. M., Jensen MP, Que L Jr. (2004). Dioxygen activation at mononuclear nonheme iron active sites: enzymes, models, and intermediates. *Chem Rev*(104), 48.

Cowtan, K. (2003). Phase Problem in X-ray Crystallography, and Its Solution. *eLS*.

D'Argenio, D. A., Gallagher, L. A., Berg, C. A., & Manoil, C. (2001). Drosophila as a model host for Pseudomonas aeruginosa infection. *J Bacteriol, 183*(4), 1466-1471.

Dann, C. E., 3rd, Bruick, R. K., & Deisenhofer, J. (2002). Structure of factor-inhibiting hypoxia-inducible factor 1: An asparaginyl hydroxylase involved in the hypoxic response pathway. *Proc Natl Acad Sci U S A, 99*(24), 15351-15356.

Dashman, T., & Kamm, J. J. (1979). Inhibition of delta-aminolevulinic acid synthetase and delta-aminolevulinic acid dehydrase by L-2-amino-4-methoxy-trans-3-butenoic acid in the rat. *Life Sci, 24*(2), 185-192.

Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., . . . Richardson, D. C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res, 35*(Web Server issue), W375-383.

Diederichs, K., & Karplus, P. A. (2013). Better models by discarding data? *Acta Crystallogr D Biol Crystallogr, 69*(Pt 7), 1215-1222.

Donald Armstrong, M. A., Dallice Mills, Bonnie Bailey, Brian Russell, Aleta Groenig, Anne Halgren, Gary Banowetz, Kerry McPhail. (2009). *Germination-Arrest Factor (GAF): 3. Determination that the herbicidal activity of GAF is associated with a ninhydrin-reactive compound and counteracted by selected amino acids* (Vol. 51): Elsevier.

Drake, E. J., & Gulick, A. M. (2008). Three-dimensional structures of Pseudomonas aeruginosa PvcA and PvcB, two proteins involved in the synthesis of 2-isocyano-6,7-dihydroxycoumarin. *J Mol Biol, 384*(1), 193-205.

Eliot, A. C., & Kirsch, J. F. (2004). Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu Rev Biochem, 73*, 383-415.

Elkins, J. M., Hewitson, K. S., McNeill, L. A., Seibel, J. F., Schlemminger, I., Pugh, C. W., . . . Schofield, C. J. (2003). Structure of factor-inhibiting hypoxia-inducible factor (HIF) reveals mechanism of oxidative modification of HIF-1 alpha. *J Biol Chem, 278*(3), 1802-1806.

Elkins, J. M., Ryle, M. J., Clifton, I. J., Dunning Hotopp, J. C., Lloyd, J. S., Burzlaff, N. I., . . . Roach, P. L. (2002). X-ray crystal structure of Escherichia coli taurine/alpha-ketoglutarate dioxygenase complexed to ferrous iron and substrates. *Biochemistry, 41*(16), 5185-5192.

Elovson, J., & Vagelos, P. R. (1968). Acyl carrier protein. X. Acyl carrier protein synthetase. *J Biol Chem, 243*(13), 3603-3611.

Emsley, P., & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr, 60*(Pt 12 Pt 1), 2126-2132.

Emsley, P., Lohkamp, B., Scott, W. G., & Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 4), 486-501.

Evans, P. R. (2011). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr D Biol Crystallogr, 67*(Pt 4), 282-292.

Evans PR, M. G. (2013). How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr*, 1204-1214.

Ewald, P. P. (1969). Introduction to the dynamical theory of X‐ray diffraction. *Introduction to the dynamical theory of X‐ray diffraction, 25*(1), 103-108.

Fernandez, L., Breidenstein, E. B., & Hancock, R. E. (2011). Creeping baselines and adaptive resistance to antibiotics. *Drug Resist Updat, 14*(1), 1-21.

Fridman, M., Balibar, C. J., Lupoli, T., Kahne, D., Walsh, C. T., & Garneau-Tsodikova, S. (2007). Chemoenzymatic formation of novel aminocoumarin antibiotics by the enzymes CouN1 and CouN7. *Biochemistry, 46*(28), 8462-8471.

Fu, X., Albermann, C., Zhang, C., & Thorson, J. S. (2005). Diversifying vancomycin via chemoenzymatic strategies. *Org Lett, 7*(8), 1513-1515.

Galm, U., Dessoy, M. A., Schmidt, J., Wessjohann, L. A., & Heide, L. (2004). In vitro and in vivo production of new aminocoumarins by a combined biochemical, genetic, and synthetic approach. *Chem Biol, 11*(2), 173-183.

Gaudelli, N. M., & Townsend, C. A. (2014). Epimerization and substrate gating by a TE domain in beta-lactam antibiotic biosynthesis. *Nat Chem Biol, 10*(4), 251-258.

Gevers, W., Kleinkauf, H., & Lipmann, F. (1969). Peptidyl transfers in gramicidin S bisoynthesis from enzyme-bound thioester intermediates. *Proc Natl Acad Sci U S A, 63*(4), 1335-1342.

Groom, C. R., Bruno, I. J., Lightfoot, M. P., & Ward, S. C. (2016). The Cambridge Structural Database. *Acta Crystallogr B Struct Sci Cryst Eng Mater, 72*(Pt 2), 171-179.

H. S. Mason, W. L. F., and E. Peterson. (1955). Oxygen transfer and electron transport by the phenolase complex. *Journal of the American Chemical Society*(77 (10)), 2.

Hahn, M., & Stachelhaus, T. (2004). Selective interaction between nonribosomal peptide synthetases is facilitated by short communication-mediating domains. *Proc Natl Acad Sci U S A, 101*(44), 15585-15590.

Hahn, M., & Stachelhaus, T. (2006). Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. *Proc Natl Acad Sci U S A, 103*(2), 275-280.

Halgren, A., Azevedo, M., Mills, D., Armstrong, D., Thimmaiah, M., McPhail, K., & Banowetz, G. (2011). Selective inhibition of Erwinia amylovora by the herbicidally active germination-arrest factor (GAF) produced by Pseudomonas bacteria. *J Appl Microbiol, 111*(4), 949-959.

Halgren, A., Maselko, M., Azevedo, M., Mills, D., Armstrong, D., & Banowetz, G. (2013). Genetics of germination-arrest factor (GAF) production by Pseudomonas fluorescens WH6: identification of a gene cluster essential for GAF biosynthesis. *Microbiology, 159*(Pt 1), 36-45.

Harris, N. C., Born, D. A., Cai, W., Huang, Y., Martin, J., Khalaf, R., . . . Zhang, W. (2018). Isonitrile Formation by a Non-Heme Iron(II)-Dependent Oxidase/Decarboxylase. *Angew Chem Int Ed Engl.*

Hashimoto, H., Pais, J. E., Zhang, X., Saleh, L., Fu, Z. Q., Dai, N., . . . Cheng, X. (2014). Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature, 506*(7488), 391-395.

Haslinger, K., Peschke, M., Brieke, C., Maximowitsch, E., & Cryle, M. J. (2015). X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. *Nature, 521*(7550), 105-109.

Hausinger, R. P. (2004). FeII/alpha-ketoglutarate-dependent hydroxylases and related enzymes. *Crit Rev Biochem Mol Biol, 39*(1), 21-68.

Hayaishi, O. (2005). An odyssey with oxygen. *Biochem Biophys Res Commun, 338*(1), 2-6.

Headd, J. J., Echols, N., Afonine, P. V., Moriarty, N. W., Gildea, R. J., & Adams, P. D. (2014). Flexible torsion-angle noncrystallographic symmetry restraints for improved macromolecular structure refinement. *Acta Crystallogr D Biol Crystallogr, 70*(Pt 5), 1346-1356.

Hegg, E. L., & Que, L., Jr. (1997). The 2-His-1-carboxylate facial triad--an emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur J Biochem, 250*(3), 625-629.

Henrick, E. K. a. K. (2005). Multiple Alignment of Protein Structures in Three Dimensions. *CompLife 2005, Springer-Verlag Berlin Heidelberg*(LNBI 3695), 2.

Hewitson, K. S., Granatino, N., Welford, R. W., McDonough, M. A., & Schofield, C. J. (2005). Oxidation by 2-oxoglutarate oxygenases: non-haem iron systems in catalysis and signalling. *Philos Trans A Math Phys Eng Sci, 363*(1829), 807-828; discussion 1035-1040.

Hidron, A. I., Edwards, J. R., Patel, J., Horan, T. C., Sievert, D. M., Pollock, D. A., . . . Participating National Healthcare Safety Network, F. (2008). NHSN annual update:

antimicrobial-resistant pathogens associated with healthcare-associated infections: annual summary of data reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2006-2007. *Infect Control Hosp Epidemiol, 29*(11), 996-1011.

Holladay, R. B. S. a. M. W. (2015). The Organic Chemistry of Drug Design and Drug Action (Third Edition). *Elsevier Inc.* .

Holm, L., & Laakso, L. M. (2016). Dali server update. *Nucleic Acids Res, 44*(W1), W351-355.

Holm, L., & Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res, 38*(Web Server issue), W545-549.

Huai, Q., Xia, Y., Chen, Y., Callahan, B., Li, N., & Ke, H. (2001). Crystal structures of 1-aminocyclopropane-1-carboxylate (ACC) synthase in complex with aminoethoxyvinylglycine and pyridoxal-5'-phosphate provide new insight into catalytic mechanisms. *J Biol Chem, 276*(41), 38210-38216.

Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., . . . Yeats, C. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Res, 37*(Database issue), D211-215.

Hutton, J. J., Jr., Trappel, A. L., & Udenfriend, S. (1966). Requirements for alpha-ketoglutarate, ferrous ion and ascorbate by collagen proline hydroxylase. *Biochem Biophys Res Commun, 24*(2), 179-184.

Icekson, I., & Apelbaum, A. (1983). Antifungal antibiotics and Siba inhibit 1-aminocyclopropane-1-carboxylic acid synthase activity. *Biochem Biophys Res Commun, 113*(2), 586-591.

Jacoby, G. A. (2009). AmpC beta-lactamases. *Clin Microbiol Rev, 22*(1), 161-182, Table of Contents.

Jaime Prilusky, C. E. F., Tzviya Zeev-Ben-Mordehai, Edwin H. Rydberg, Orna Man, Jacques S. Beckmann. Israel Silman, Joel L. Sussman. (2005). FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded *Bioinformatics, 21*(16), 3435-3438.

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol, 292*(2), 195-202.

K. Xiong, J. J. F. (1996). *Comparison of rhizobitoxine-induced inhibition of β-cystathionase from different bradyrhizobia and soybean genotypes* (Vol. 186): Kluwer Academic Publishers.

Kabsch, W. (2010). Xds. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 2), 125-132.

Kantardjieff, K. A., & Rupp, B. (2003). Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Sci, 12*(9), 1865-1871.

Keating, T. A., Marshall, C. G., Walsh, C. T., & Keating, A. E. (2002). The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat Struct Biol, 9*(7), 522-526.

Keating, T. A., Miller, D. A., & Walsh, C. T. (2000). Expression, purification, and characterization of HMWP2, a 229 kDa, six domain protein subunit of Yersiniabactin synthetase. *Biochemistry, 39*(16), 4729-4739.

Keegan, R. M., & Winn, M. D. (2008). MrBUMP: an automated pipeline for molecular replacement. *Acta Crystallogr D Biol Crystallogr, 64*(Pt 1), 119-124.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc, 10*(6), 845-858.

Kleinkauf, H., Gevers, W., & Lipmann, F. (1969). Interrelation between activation and polymerization in gramicidin S biosynthesis. *Proc Natl Acad Sci U S A, 62*(1), 226-233.

Kleinkauf, H., & Von Dohren, H. (1996). A nonribosomal system of peptide biosynthesis. *Eur J Biochem, 236*(2), 335-351.

Krissinel, E. (2010). Crystal contacts as nature's docking solutions. *journal of computational chemistry, 31*(1), 133-143.

Krissinel, E., & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr, 60*(Pt 12 Pt 1), 2256-2268.

Krissinel, E., & Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J Mol Biol, 372*(3), 774-797.

Laskowski, R. A., Jablonska, J., Pravda, L., Varekova, R. S., & Thornton, J. M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Sci, 27*(1), 129-134.

Lee, J., Wu, J., Deng, Y., Wang, J., Wang, C., Wang, J., . . . Zhang, L. H. (2013). A cell-cell communication signal integrates quorum sensing and stress response. *Nat Chem Biol, 9*(5), 339-343.

Lee, X., Azevedo, M. D., Armstrong, D. J., Banowetz, G. M., & Reimmann, C. (2013). The Pseudomonas aeruginosa antimetabolite L-2-amino-4-methoxy-trans-3-butenoic acid inhibits growth of Erwinia amylovora and acts as a seed germination-arrest factor. *Environ Microbiol Rep, 5*(1), 83-89.

Lee, X., Fox, A., Sufrin, J., Henry, H., Majcherczyk, P., Haas, D., & Reimmann, C. (2010). Identification of the biosynthetic gene cluster for the Pseudomonas aeruginosa antimetabolite L-2-amino-4-methoxy-trans-3-butenoic acid. *J Bacteriol, 192*(16), 4251-4255.

Lee, X., Reimmann, C., Greub, G., Sufrin, J., & Croxatto, A. (2012). The Pseudomonas aeruginosa toxin L-2-amino-4-methoxy-trans-3-butenoic acid inhibits growth and induces encystment in Acanthamoeba castellanii. *Microbes Infect, 14*(3), 268-272.

Levin, E. J., Kondrashov, D. A., Wesenberg, G. E., & Phillips, G. N., Jr. (2007). Ensemble refinement of protein crystal structures: validation and application. *Structure, 15*(9), 1040-1052.

Li, M. Z., & Elledge, S. J. (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Methods, 4*(3), 251-256.

Lloyd, A. G. P. a. M. D. (2000). The iron(II) and 2-oxoacid-dependent dioxygenases and their role in metabolism. *Nat. Prod. Rep.*(17), 17.

Lloyd, M. D., Lee, H. J., Harlos, K., Zhang, Z. H., Baldwin, J. E., Schofield, C. J., . . . Bhikhabhai, R. (1999). Studies on the active site of deacetoxycephalosporin C synthase. *J Mol Biol, 287*(5), 943-960.

Long, F., Vagin, A. A., Young, P., & Murshudov, G. N. (2008). BALBES: a molecular-replacement pipeline. *Acta Crystallogr D Biol Crystallogr, 64*(Pt 1), 125-132.

Losey, H. C., Jiang, J., Biggins, J. B., Oberthur, M., Ye, X. Y., Dong, S. D., . . . Walsh, C. T. (2002). Incorporation of glucose analogs by GtfE and GtfD from the vancomycin biosynthetic pathway to generate variant glycopeptides. *Chem Biol, 9*(12), 1305-1314.

Lowell D. Owens, J. F. T., R. G. Pitcher and T. Williams. (1972). Structure of rhizobitoxine, an antimetabolic enol-ether amino-acid from Rhizobium japonicum. *J. Chem. Soc., Chem. Commun.*, 714-714.

Lyczak, J. B., Cannon, C. L., & Pier, G. B. (2000). Establishment of Pseudomonas aeruginosa infection: lessons from a versatile opportunist. *Microbes Infect, 2*(9), 1051-1060.

M.A. Larkin, G. B., N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez J.D. Thompson, T.J. Gibson, D.G. Higgins. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics, 23*(21), 2947-2948.

Mahajan-Miklos, S., Tan, M. W., Rahme, L. G., & Ausubel, F. M. (1999). Molecular mechanisms of bacterial virulence elucidated using a Pseudomonas aeruginosa-Caenorhabditis elegans pathogenesis model. *Cell, 96*(1), 47-56.

Marahiel, M. A., Stachelhaus, T., & Mootz, H. D. (1997). Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem Rev, 97*(7), 2651-2674.

Matthews, B. W. (1968). Solvent content of protein crystals. *J Mol Biol, 33*(2), 491-497.

Mattoo, A. K., Anderson, J. D., Chalutz, E., & Lieberman, M. (1979). Influence of enol ether amino acids, inhibitors of ethylene biosynthesis, on aminoacyl transfer RNA synthetases and protein synthesis. *Plant Physiol, 64*(2), 289-292.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., & Read, R. J. (2007). Phaser crystallographic software. *J Appl Crystallogr, 40*(Pt 4), 658-674.

Miles, E. W. (1975). A new type of pyridoxal-P enzyme catalyzed reaction: the conversion of beta, gamma-unsaturated amino acids to saturated alpha-keto acids by tryptophan synthase. *Biochem Biophys Res Commun, 66*(1), 94-102.

Miyata, S., Casey, M., Frank, D. W., Ausubel, F. M., & Drenkard, E. (2003). Use of the Galleria mellonella caterpillar as a model host to study the role of the type III secretion system in Pseudomonas aeruginosa pathogenesis. *Infect Immun, 71*(5), 2404-2413.

Mount, D. W. (2007). Using the Basic Local Alignment Search Tool (BLAST). *CSH Protoc, 2007*, pdb top17.

Mueller-Dieckmann, C., Bowler, M. W., Carpentier, P., Flot, D., McCarthy, A. A., Nanao, M. H., Nurizzo, D., Pernot, P., Popov, A., Round, A., Royant, A., de Sanctis, D., von

Stetten, D., Leonard, G. A. . (2015). The status of the macromolecular crystallography beamlines at the European Synchrotron Radiation Facility. *Eur. Phys. J. Plus 130*, 1.

Muller, I., Kahnert, A., Pape, T., Sheldrick, G. M., Meyer-Klaucke, W., Dierks, T., . . . Uson, I. (2004). Crystal structure of the alkylsulfatase AtsK: insights into the catalytic mechanism of the Fe(II) alpha-ketoglutarate-dependent dioxygenase superfamily. *Biochemistry, 43*(11), 3075-3088.

N. A. de Bruyne, R. H. (1952). Adhesion and Adhesives. *Science, 115*(2995), 2.

O'Brien, J. R., Schuller, D. J., Yang, V. S., Dillard, B. D., & Lanzilotta, W. N. (2003). Substrate-induced conformational changes in Escherichia coli taurine/alpha-ketoglutarate dioxygenase and insight into the oligomeric structure. *Biochemistry, 42*(19), 5547-5554.

Osamu Hayaishi, M. K., and Simon Rothberg. (1955). Mechanism of the pyrocatechase reaction. *J. Am. Chem. Soc.*(77 (20)), 2.

Percudani, R., & Peracchi, A. (2003). A genomic overview of pyridoxal-phosphate-dependent enzymes. *EMBO Rep, 4*(9), 850-854.

Peters, G. J. (2014). Novel developments in the use of antimetabolites. *Nucleosides, Nucleotides Nucleic Acids, 33*(4-6), 17.

Peters, G. J., van der Wilt, C. L., van Moorsel, C. J., Kroep, J. R., Bergman, A. M., & Ackland, S. P. (2000). Basis for effective combination cancer chemotherapy with antimetabolites. *Pharmacol Ther, 87*(2-3), 227-253.

Poole, K. (2004). Efflux-mediated multiresistance in Gram-negative bacteria. *Clin Microbiol Infect, 10*(1), 12-26.

Pruess, D. L., Scannell, J. P., Kellett, M., Ax, H. A., Janecek, J., Williams, T. H., . . . Berger, J. (1974). Antimetabolites produced by microorganisms. X. L-2-amino-4-(2-aminoethoxy)-trans-3-butenoic acid. *J Antibiot (Tokyo), 27*(4), 229-233.

Puk, O., Bischoff, D., Kittel, C., Pelzer, S., Weist, S., Stegmann, E., . . . Wohlleben, W. (2004). Biosynthesis of chloro-beta-hydroxytyrosine, a nonproteinogenic amino acid of the peptidic backbone of glycopeptide antibiotics. *J Bacteriol, 186*(18), 6093-6100.

Puk, O., Huber, P., Bischoff, D., Recktenwald, J., Jung, G., Sussmuth, R. D., . . . Pelzer, S. (2002). Glycopeptide biosynthesis in Amycolatopsis mediterranei DSM5908: function of a halogenase and a haloperoxidase/perhydrolase. *Chem Biol, 9*(2), 225-235.

Purpero, V., & Moran, G. R. (2007). The diverse and pervasive chemistries of the alpha-keto acid dependent enzymes. *J Biol Inorg Chem, 12*(5), 587-601.

Que, L., Jr., & Ho, R. Y. (1996). Dioxygen Activation by Enzymes with Mononuclear Non-Heme Iron Active Sites. *Chem Rev, 96*(7), 2607-2624.

Rahme, L. G., Ausubel, F. M., Cao, H., Drenkard, E., Goumnerov, B. C., Lau, G. W., . . . Tompkins, R. G. (2000). Plants and animals share functionally common bacterial virulence factors. *Proc Natl Acad Sci U S A, 97*(16), 8815-8821.

Rahme, L. G., Stevens, E. J., Wolfort, S. F., Shao, J., Tompkins, R. G., & Ausubel, F. M. (1995). Common virulence factors for bacterial pathogenicity in plants and animals. *Science, 268*(5219), 1899-1902.

Rahme, L. G., Tan, M. W., Le, L., Wong, S. M., Tompkins, R. G., Calderwood, S. B., & Ausubel, F. M. (1997). Use of model plant hosts to identify Pseudomonas aeruginosa virulence factors. *Proc Natl Acad Sci U S A, 94*(24), 13245-13250.

Rando, R. R. (1974). Beta, gamma unsaturated amino acids as irreversible enzyme inhibitors. *Nature, 250*(467), 586-587.

Rando, R. R., Relyea, N., & Cheng, L. (1976). Mechanism of the irreversible inhibition of aspartate aminotransferase by the bacterial toxin L-2-amino-4-methoxy-trans-3-butenoic acid. *J Biol Chem, 251*(11), 3306-3312.

Rausch, C., Hoof, I., Weber, T., Wohlleben, W., & Huson, D. H. (2007). Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol Biol, 7*, 78.

Roach, P. L., Clifton, I. J., Fulop, V., Harlos, K., Barton, G. J., Hajdu, J., . . . Baldwin, J. E. (1995). Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes. *Nature, 375*(6533), 700-704.

Robert, X., & Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res, 42*(Web Server issue), W320-324.

Robin E.Mitchell, J. M. C. (1991). Biosynthetic pathway to rhizobitoxine in Pseudomonas andropogonis. *Phytochemistry, 30*(6), 1809-1814.

Rojas Murcia, N., Lee, X., Waridel, P., Maspoli, A., Imker, H. J., Chai, T., . . . Reimmann, C. (2015). The Pseudomonas aeruginosa antimetabolite L -2-amino-4-methoxy-trans-3-butenoic acid (AMB) is made from glutamate and two alanine residues via a thiotemplate-linked tripeptide precursor. *Front Microbiol, 6*, 170.

Roskoski, R., Jr., Gevers, W., Kleinkauf, H., & Lipmann, F. (1970). Tyrocidine biosynthesis by three complementary fractions from Bacillus brevis (ATCC 8185). *Biochemistry, 9*(25), 4839-4845.

Roskoski, R., Jr., Kleinkauf, H., Gevers, W., & Lipmann, F. (1970). Isolation of enzyme-bound peptide intermediates in tyrocidine biosynthesis. *Biochemistry, 9*(25), 4846-4851.

Rupp, B. (2009). *Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology*: Garland Science

S. Brennan, P. L. C. (1992). A suite of programs for calculating x‐ray absorption, reflection, and diffraction performance for a variety of materials at arbitrary wavelengths. *Review of Scientific Instruments, 63*(850).

Sahm, U., Knobloch, G., & Wagner, F. (1973). Isolation and characterization of the methionine antagonist L-2-amino-4-methoxy-trans-3-butenoic acid from Pseudomonas aeruginosa grown on n-paraffin. *J Antibiot (Tokyo), 26*(7), 389-390.

Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A., & Essen, L. O. (2007). Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure, 15*(7), 781-792.

Scannel, J. P., Pruess, D. L., Demny, T. C., Sello, L. H., & Williams, T. (1972). Antimetabolites produced by microorganisms. V. L-2-amino-4-methoxy-trans-3-butenoic acid. *J Antibiot (Tokyo), 25*(2), 122-127.

Schobert, M., & Jahn, D. (2010). Anaerobic physiology of Pseudomonas aeruginosa in the cystic fibrosis lung. *Int J Med Microbiol, 300*(8), 549-556.

Schofield, C. J., & Zhang, Z. (1999). Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr Opin Struct Biol, 9*(6), 722-731.

Schrödinger. (2009). The AxPyMOL Molecular Graphics Plugin for PowerPoint, Version 2.0. *LLC.*

Schweizer, H. P. (2003). Efflux as a mechanism of resistance to antimicrobials in Pseudomonas aeruginosa and related bacteria: unanswered questions. *Genet Mol Res, 2*(1), 48-62.

Shaan L. Gellatly, R. E. W. H. (2013). *Pseudomonas aeruginosa* : new insights into pathogenesis and host defenses. *Pathogens and Disease, 67*(3), 15.

Simon, R. H., Scoggin, C. H., & Patterson, D. (1981). Hydrogen peroxide causes the fatal injury to human fibroblasts exposed to oxygen radicals. *J Biol Chem, 256*(14), 7181-7186.

Smith, S. B., & Freedland, R. A. (1981). Functional inhibition of cytosolic and mitochondrial aspartate aminotransferase by L-2-amino-4-methoxy-trans-3-butenoic acid in isolated rat hepatocytes and mitochondria. *Arch Biochem Biophys, 209*(2), 335-341.

Smith, T. J. (1995). MolView: a program for analyzing and displaying atomic structures on the Macintosh personal computer. *J Mol Graph, 13*(2), 122-125, 115.

Stachelhaus, T., Mootz, H. D., Bergendahl, V., & Marahiel, M. A. (1998). Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J Biol Chem, 273*(35), 22773-22781.

Starkey, M., & Rahme, L. G. (2009). Modeling Pseudomonas aeruginosa pathogenesis in plant hosts. *Nat Protoc, 4*(2), 117-124.

Stoll, E., Froyshov, O., Holm, H., Zimmer, T. L., & Laland, S. G. (1970). On the mechanism of gramicidin S formation from intermediate peptides. *FEBS Lett, 11*(5), 348-352.

Terwilliger, T. C. (2003). Improving macromolecular atomic models at moderate resolution by automated iterative model building, statistical density modification and refinement. *Acta Crystallogr D Biol Crystallogr, 59*(Pt 7), 1174-1182.

Terwilliger, T. C., Adams, P. D., Read, R. J., McCoy, A. J., Moriarty, N. W., Grosse-Kunstleve, R. W., . . . Hung, L. W. (2009). Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr, 65*(Pt 6), 582-601.

Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L. W., . . . Adams, P. D. (2008). Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr D Biol Crystallogr, 64*(Pt 1), 61-69.

Tillett, D., Dittmann, E., Erhard, M., von Dohren, H., Borner, T., & Neilan, B. A. (2000). Structural organization of microcystin biosynthesis in Microcystis aeruginosa PCC7806: an integrated peptide-polyketide synthetase system. *Chem Biol, 7*(10), 753-764.

Tisdale, M. J. (1980). The effect of the methionine antagonist L-2-amino-4-methoxy-trans-3-butenoic acid on the growth and metabolism of Walker carcinosarcoma in vitro. *Biochem Pharmacol, 29*(4), 501-508.

Valko, M., Rhodes, C. J., Moncol, J., Izakovic, M., & Mazur, M. (2006). Free radicals, metals and antioxidants in oxidative stress-induced cancer. *Chem Biol Interact, 160*(1), 1-40.

Van de Poel, B., & Van Der Straeten, D. (2014). 1-aminocyclopropane-1-carboxylic acid (ACC) in plants: more than just the precursor of ethylene! *Front Plant Sci, 5*, 640.

van Ditmarsch, D., Boyle, K. E., Sakhtah, H., Oyler, J. E., Nadell, C. D., Deziel, E., . . . Xavier, J. B. (2013). Convergent evolution of hyperswarming leads to impaired biofilm formation in pathogenic bacteria. *Cell Rep, 4*(4), 697-708.

van Wageningen, A. M., Kirkpatrick, P. N., Williams, D. H., Harris, B. R., Kershaw, J. K., Lennard, N. J., . . . Solenberg, P. J. (1998). Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem Biol, 5*(3), 155-162.

Velkov, T., & Lawen, A. (2003). Mapping and molecular modeling of S-adenosyl-L-methionine binding sites in N-methyltransferase domains of the multifunctional polypeptide cyclosporin synthetase. *J Biol Chem, 278*(2), 1137-1148.

Vining, L. C. (1990). Functions of secondary metabolites. *Annu Rev Microbiol, 44*, 395-427.

Walden, H. (2010). Selenium incorporation using recombinant techniques. *Acta Crystallogr D Biol Crystallogr, 66*(Pt 4), 352-357.

Walker, T. S., Bais, H. P., Deziel, E., Schweizer, H. P., Rahme, L. G., Fall, R., & Vivanco, J. M. (2004). Pseudomonas aeruginosa-plant root interactions. Pathogenicity, biofilm formation, and root exudation. *Plant Physiol, 134*(1), 320-331.

Waltersperger, S., Olieric, V., Pradervand, C., Glettig, W., Salathe, M., Fuchs, M. R., . . . Wang, M. (2015). PRIGo: a new multi-axis goniometer for macromolecular crystallography. *J Synchrotron Radiat, 22*(4), 895-900.

Williams, B. J., Dehnbostel, J., & Blackwell, T. S. (2010). Pseudomonas aeruginosa: host defence in lung diseases. *Respirology, 15*(7), 1037-1056.

Wilmouth, R. C., Turnbull, J. J., Welford, R. W., Clifton, I. J., Prescott, A. G., & Schofield, C. J. (2002). Structure and mechanism of anthocyanidin synthase from Arabidopsis thaliana. *Structure, 10*(1), 93-103.

Winsor, G. L., Griffiths, E. J., Lo, R., Dhillon, B. K., Shay, J. A., & Brinkman, F. S. (2016). Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database. *Nucleic Acids Res, 44*(D1), D646-653.

Xu, H., Heide, L., & Li, S. M. (2004). New aminocoumarin antibiotics formed by a combined mutational and chemoenzymatic approach utilizing the carbamoyltransferase NovN. *Chem Biol, 11*(5), 655-662.

Yasuta, T., Satoh, S., & Minamisawa, K. (1999). New assay for rhizobitoxine based on inhibition of 1-aminocyclopropane-1-carboxylate synthase. *Appl Environ Microbiol, 65*(2), 849-852.

Yates, S. P., & Merrill, A. R. (2004). Elucidation of eukaryotic elongation factor-2 contact sites within the catalytic domain of Pseudomonas aeruginosa exotoxin A. *Biochem J, 379*(Pt 3), 563-572.

Zhang, Y. F., Han, K., Chandler, C. E., Tjaden, B., Ernst, R. K., & Lory, S. (2017). Probing the sRNA regulatory landscape of P. aeruginosa: post-transcriptional control of determinants of pathogenicity and antibiotic susceptibility. *Mol Microbiol, 106*(6), 919-937.

Zhang, Z., Ren, J., Stammers, D. K., Baldwin, J. E., Harlos, K., & Schofield, C. J. (2000). Structural origins of the selectivity of the trifunctional oxygenase clavaminic acid synthase. *Nat Struct Biol, 7*(2), 127-133.

Zheng, H., Chordia, M. D., Cooper, D. R., Chruszcz, M., Muller, P., Sheldrick, G. M., & Minor, W. (2014). Validation of metal-binding sites in macromolecular structures with the CheckMyMetal web server. *Nat Protoc, 9*(1), 156-170.

Zheng, H., Cooper, D. R., Porebski, P. J., Shabalin, I. G., Handing, K. B., & Minor, W. (2017). CheckMyMetal: a macromolecular metal-binding validation tool. *Acta Crystallogr D Struct Biol, 73*(Pt 3), 223-233.

Zhou, J., Kelly, W. L., Bachmann, B. O., Gunsior, M., Townsend, C. A., & Solomon, E. I. (2001). Spectroscopic studies of substrate interactions with clavaminate synthase 2, a multifunctional alpha-KG-dependent non-heme iron enzyme: correlation with mechanisms and reactivities. *J Am Chem Soc, 123*(30), 7388-7398.

Zhu, W., Arceneaux, J. E., Beggs, M. L., Byers, B. R., Eisenach, K. D., & Lundrigan, M. D. (1998). Exochelin genes in Mycobacterium smegmatis: identification of an ABC transporter and two non-ribosomal peptide synthetase genes. *Mol Microbiol, 29*(2), 629-639.

Acknowledgements

During the four years study at Structure and Function of Proteins (SFPR) group supervised by Prof. Dr. Wulf Blankenfeldt in Helmholtz Centre for Infection Research (HZI), I have been always supported by my mentor, colleagues and friends. Here I want to deliver my most sincere thanks to all of you.

Prof. Dr. Wulf Blankenfeldt, as my PhD mentor, for providing me the chance to work here and the interesting project, for the academic guides and suggestions that helped me a lot, for his encourage that made me confident.

Dr. Jan Pippel for the instruction of gas chromatography mass spectrometry; **Dr. Konrad Büssow** for the plasmids construction; **Dr. Peer Lukat** for the suggestions of generating figures with PyMOL; **Kristin Althof** for carrying the side project of methyltransferase domain of AmbE.

Prof. Dr. Cornelia Reimmann and **Dr. Stephan Heeb** for providing the constructs of AMB synthesis related proteins; **Prof. Chris Calderone** for the hypothesis of AMB synthesis; **Prof. Dr. Rolf Müller** and **Dr. David Auerbach** for providing the protocol of off-loading experiments.

Prof. Dr. Susanne Häußler and **Prof. Dr Marc Stadler**, as the members of my thesis committee, for the scientific discussions and advice for my project.

My dear colleagues from SFPR group: **Monika Popp, Nina Schweimmlein, Florian Witzgall, Maurice Diwo, FangFang Chen, Pooja Sadana, Thomas Klünemann, Juliane poppe, Stefan Schmelz, Tobias Bock, Christian Feiler, Allegra Vit, Christina Diederich, Claudia Hanko, Ute Widow, Kwanghoon Sung** and other **SFPRers** and **SBAUers** for constant and friendly help on both academic work and German style life, for their discussions and suggestions contributing to a wonderful environment in the lab.

All my lovely Chinese friends in HZI and Braunschweig for enriching our traditional life.

The **Graduate School of Helmholtz Centre for Infection Research** for providing academic communications through organizing PhD symposiums and annual retreats.

The **PETRA III** (DESY, EMBL, Hamburg), **ESRF** (Grenoble, France) and **SLS** (PSI, Villigen, Swiss) for datasets collection and scientific instructions.

China Scholarship Council for the financial support during my PhD study.

Thank **Ying An** for making my life colorful.

Last but not least, thank my parents and sisters for their constant understanding and support. I cannot make this without them. Wish you all the best!

Thank you

Danke

谢谢

Curriculum Vitae

Personal information

Name	Yafei Xiao
Date of birth	06 June, 1988
Place of birth	Henan, China
Nationality	Chinese

Work experience and Education

September 2014 – present

Structure and Function of Proteins, Helmholtz Centre for Infection Research, Braunschweig.

Ph. D. candidate under supervision of Prof. Dr. Wulf Blankenfeldt

September 2011 – March 2014

State Key Laboratory of Microbial metabolism, Shanghai Jiao Tong University, Shanghai.

Master of Science under supervision of Prof. Dr. Geng Wu

September 2011 – July 2013

ZHIYUAN College, Shanghai Jiao Tong University, Shanghai.

Teaching Assistant for Prof. Dr. Jianjun Chu

September 2007 – June 2011

School of life science and Technology, China Pharmaceutical University, Nanjing.

Bachelor of Science under supervision of Prof. Dr. Yu Ou

Scholarship

September 2014 – present

China Scholarship Council (CSC) scholarship from China government.

Publications

Zhenyi Zhang,*, Senem Akyildiz,*, Yafei Xiao,*, Zhongchao Gai,*, Ying An, Jürgen Behrens, Geng Wu. Structures of the APC–ARM domain in complexes with discrete Amer1/WTX fragments reveal that it uses a consensus mode to recognize its binding partners. *Cell Discovery* (2015) 1, 15016.

Yafei Xiao, Geng Wu. Identification of a New Binding Site in Amer1 (365-375) Associated with APC-ARM Domain. *Chinese Journal of Biochemistry and Molecular Biol*, 2014, 30(4): 383-389.